PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/82, 9/10, 1/21, 15/62, A01H 5/00

A2

(11) International Publication Number: WO 98/50570

(43) International Publication Date: 12 November 1998 (12.11.98)

(21) International Application Number: PCT/US98/09522

(22) International Filing Date: 8 May 1998 (08.05.98)

(30) Priority Data:
60/046,857 8 May 1997 (08.05.97) US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application
US 60/046,857 (CIP)
Filed on 8 May 1997 (08.05.97)

(71) Applicant *(for all designated States except US)*: THE REGENTS OF THE UNIVERSITY OF MICHIGAN [US/US]; Room 2071, Wolverine Tower, 3003 South State Street, Ann Arbor, MI 48109–1280 (US).

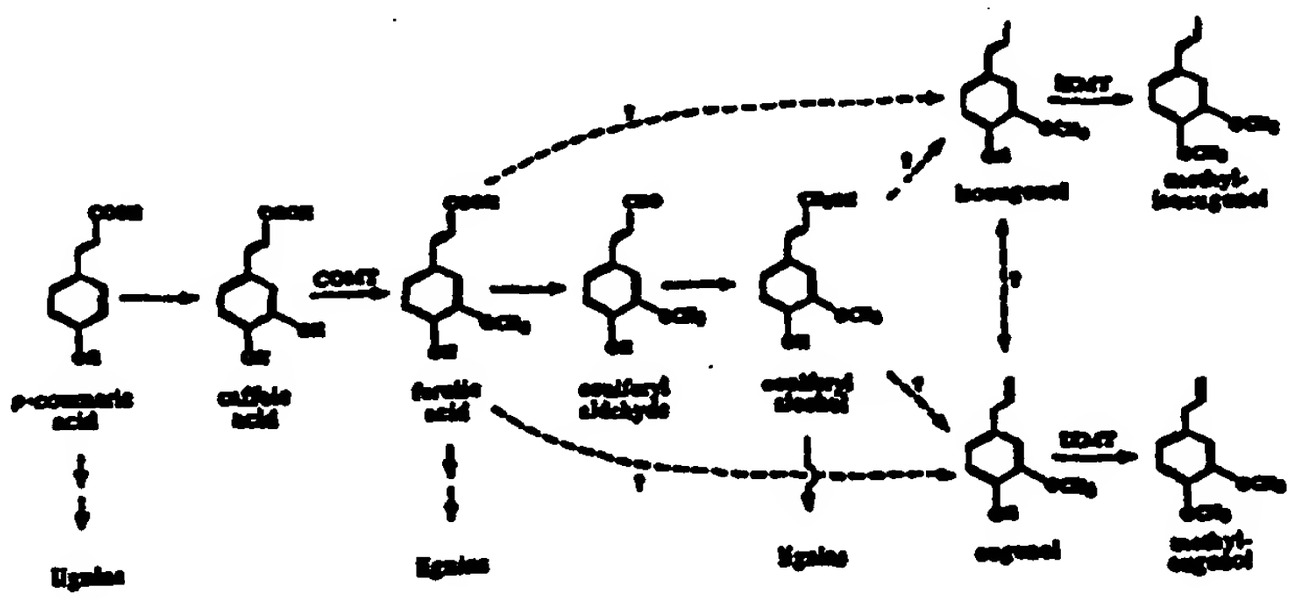
(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: PICHERSKY, Eran [US/US]; 325 Burlington Woods Drive, No. 180, Ann Arbor, MI 48104 (US). WANG, Jihong [CN/US]; Apartment 21, 1738 Murfin, Ann Arbor, MI 48105 (US).

(74) Agent: MCMILLIAN, Nabeela, R.; Arnold, White & Durkee, P.O. Box 4433, Houston, TX 77210 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: METHODS AND COMPOSITIONS FOR USE OF (ISO)EUGENOL METHYLTRANSFERASE



(57) Abstract

The present invention concerns DNA compositions and their use in manipulating the biosynthesis of compound in plants. More specifically, the present invention has identified a novel gene and protein involved in the phenylpropanoid pathways. *S*–adenosyl–L–Met:(iso)eugenol *O*–methyltransferas (IEMT) catalyzes a methyl transfer to the C4 hydroxyl position of compounds of the phenylpropanoid biosynthetic pathway. Methods and compositions for the use thereof are disclosed herein.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

DESCRIPTION

METHODS AND COMPOSITIONS FOR USE OF (ISO)EUGENOL METHYLTRANSFERASE

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates generally to the fields of plant molecular biology, biochemistry plant physiology. More particularly, it concerns DNA compositions and their use in manipulating the biosynthesis of compound in plants.

2. Description of Related Art

A. Plant Flavor and Scent

Phenylpropanoids are often used in many perfumes and food seasonings. Eugenol and isoeugenol have a pungent clove-like aroma, whereas methyleugenol and methylisoeugenol have a soft herbaceous smell, reminiscent of freshly cut leaves. These four compounds alone and in combination contribute to the distinctive flavor of several herbs and fruits (*e.g.*, basil, tomato), and they are also important components of the floral scent of various species, including *Clarkia breweri* (Raguso and Pichersky, 1995).

With the advances made in molecular biology it is now possible to genetically engineer favorable traits such as those mentioned above into plants to produce crop plants which can be selected for flavor and scent.

There has been much speculation on the biochemical pathways of phenylpropanoid production. However, despite the importance of floral scent to plant reproduction and evolution and flavor to the desirability of a plant product, the biochemical and genetic basis of scent

production has received little attention. Previous reports have failed to identify and purify specific enzymes involved in the biosynthesis of scent components in flowers. Since many scent components also are found in floral tissues in bound, non-volatile forms such as glycosides, it was originally hypothesized that scent compounds could possibly be synthesized elsewhere in the plant, bound into glycosides, and then transported to the emitting part of the flower, where they could be broken down to release the volatile components (Ackermann *et al.*, 1989; Watanabe *et al.*, 1993). However, direct and reproducible evidence of the transport of free scent constituents or their glycosides from vegetative tissue to floral tissue is lacking.

The inventors previously have investigated the biosynthetic pathways of scent components and their location in the plant by examining *Clarkia breweri* (Gray) Greene (Onagraceae), an annual plant native to California. The strong, sweet fragrance of *C. breweri* consists of 8 to 12 different volatiles that fall into two groups: monoterpenoids and phenylpropanoids-benzenoids (Raguso and Pichersky, 1995). The inventors have previously shown that linalool, an acyclic monoterpene, is synthesized in *C. breweri* petals and other floral tissues in a reaction catalyzed by *S*-linalool synthase (LIS) (Pichersky *et al.*, 1994; Pichersky *et al.*, 1995), and that the enzyme activity is regulated at a pretranslational level (Dudareva *et al.*, 1996). A closely related species, *Clarkia concinna* (from which *C. breweri* is believed to have evolved) (Raguso and Pichersky, 1995), also possesses the gene encoding LIS, but the expression of this gene in this nonscented species is limited to the stigma, and the level of expression is much lower than in the stigma of *C. breweri*. Thus, linalool production in *C. breweri* flowers involves a change in regulation of an existing gene.

Elucidation of the phenylpropanoid biosynthetic pathway, as well as the cloning of genes involved in the pathway, would allow for the production of transgenic plants with enhanced profiles of phenylpropanoid biosynthesis. Such plants would have improved flavor and fragrance and represent a significant advance to agriculture. To date, however, reaching the goal of producing such plants has been severely limited by the general lack of information regarding the phenylpropanoid biosynthetic pathway and the genes which encode enzymes in the pathway.

B. Lignin Production

A further aspect of phenylpropanoid biosynthetic pathway is that it bifurcates from the lignin biosynthetic pathway and is, therefore, involved in plant lignification. Digestibility of crops is determined, among other factors, by the amount of lignification which has taken place during growth of the plants and the degree of secondary modification of lignin deposited. Beside cellulose and other polysaccharides, lignins are an essential component of the cell wall in tissues like the sclerenchyma and the xylem of vascular plants. They play an important role in the conducting function of the xylem by reducing the permeability of the cell wall to water. They are also responsible for the rigidity of the cell wall, and, in woody tissues, they act as a bonding agent between cells, imparting to the plant a resistance towards impact, compression and bending. Finally, they are involved in mechanisms of resistance to pathogens by impeding the penetration or the propagation of the pathogenic agent.

Thus, plants with a reduced amount of lignin or modified lignin composition would be more efficiently used as crop plants and as fodder plants. Furthermore, lignin may have a negative effect on plant growth.

Thus, a reduction of the lignification in crops such as tomatoes, wheat, oilseed rape, sugar beet or maize might presumably increase the grain yield. Trees with reduced lignin contents or altered lignin structure will lead to a reduction in the cost of the paper as less lignin will have to be removed during the pulping process. On the other hand, novel papers may be produced due to the purity of cellulose fiber which could otherwise not be produced.

SUMMARY OF THE INVENTION

Thus it is an objective of the present invention to provide DNA compositions and their use in manipulating the biosynthesis of lignin and flavoring compounds in plants. More specifically, the present invention identifies a novel gene and protein involved in the

phenylpropanoid pathways. Methods and compositions for the use of IEMT are disclosed herein.

Thus, in order to achieve the objectives of the present invention there is provided an isolated nucleic acid comprising a nucleic acid segment coding for (iso)eugenol methyl transferase (IEMT), or an active fragment thereof. The nucleic acid segment may encode full length IEMT. In certain preferred embodiments, the IEMT has the sequence of SEQ ID NO:2. In other preferred embodiments, the nucleic acid segment has the sequence of SEQ ID NO:1.

The present invention further provides an expression vector comprising a nucleic acid segment coding for (iso)eugenol methyl transferase (IEMT) and a promoter operatively linked to the nucleic acid segment. In particular embodiments, the promoter may be selected from the group consisting of 35S promoter and CHS-A promoter. In other embodiments, the promoter may be a fruit-specific promoter or a leaf-specific promoter. In certain embodiments, the expression vector may further comprise an origin of replication and a polyadenylation signal.

The present invention also contemplates a method for increasing the synthesis of methyl-isoeugenol in a plant cell comprising the steps of providing plant cells; contacting the plant cell with a nucleic acid segment coding for (iso)eugenol methyl transferase (IEMT) and a promoter operatively linked to the nucleic acid segment; and selecting a cell having the nucleic acid stably integrated into is genome. In certain embodiments, the plant cell may be a monocot, and in other embodiments, the plant cell is a dicot. In a particularly preferred embodiment, the dicot plant cell is a tomato cell. In some aspects of the present invention, the contacting comprises microprojectile bombardment, electroporation or *Agrobacterium*-mediated transformation. In other aspects, the selecting comprises identifying a cell having drug resistance. In those aspects in which the selecting comprises drug resistance, the drug resistance utilized may be selected from the group consisting of kanamycin- and hygromycin-resistance.

The present invention further provides a transgenic plant cell having, incorporated into its genome, a nucleic acid segment coding for (iso)eugenol methyl transferase (IEMT) and a promoter operatively linked to the nucleic acid segment. The plant cell may be a monocot or a dicot. In a preferred embodiment, the transgenic plant cell is a tomato cell.

Other aspects of the invention provide a transgenic plant having, incorporated into the genome of cells of the plant, a nucleic acid segment coding for (iso)eugenol methyl transferase (IEMT) and a promoter operatively linked to the nucleic acid segment.

Also provided by the present invention is a method for decreasing the (iso)eugenol content of a plant cell comprising the steps of providing plant cells; contacting the plant cell with a nucleic acid segment coding for (iso)eugenol methyl transferase (IEMT) and a promoter operatively linked to the nucleic acid segment; and selecting a cell having the nucleic acid stably integrated into its genome. In preferred embodiments, the contacting comprises microprojectile bombardment, electroporation or Agrobacterium-mediated transformation.

The present invention further contemplates a method for decreasing lignin biosynthesis in a plant cell comprising the steps of providing a plant cell; contacting the plant cell with nucleic acid segment encoding a chimeric enzyme having amino acid sequences from (iso)eugenol methyl transferase (IEMT) and coumaric acid methyl transferase (COMT), wherein the substrate specificity of the enzyme includes *p*-coumaric acid, caffeic acid, ferulic acid, eugenol and isoeugenol and the activity comprises 4-hydroxyl methylation; and selecting a cell having the nucleic acid segment stably integrated into its genome.

In yet another aspect, the present invention provides a transgenic plant, having incorporated into the genome of cells of the plant, a nucleic acid segment encoding a chimeric enzyme having amino acid sequences from (iso)eugenol methyl transferase (IEMT) and caffeic acid methyl transferase (COMT), wherein the substrate specificity of the enzyme includes *p*-coumaric acid, caffeic acid, ferulic acid, eugenol and isoeugenol and the activity comprises 4-hydroxyl methylation and a promoter operatively linked to the nucleic acid segment.

In still a further aspect the present invention provides an isolated nucleic acid segment encoding a chimeric enzyme having amino acid sequences from (iso)eugenol methyl transferase (IEMT) and caffeic acid methyl transferase (COMT), wherein the substrate specificity of the enzyme includes *p*-coumaric acid, caffeic acid, ferulic acid, eugenol and isoeugenol and the activity comprises 4-hydroxyl methylation and a promoter operatively linked to the nucleic acid segment.

Other objects, features and advantages of the present invention will become apparent from the following detailed description. It should be understood, however, that the detailed description and the specific examples, while indicating preferred embodiments of the invention, are given by way of illustration only, since various changes and modifications within the spirit and scope of the invention will become apparent to those skilled in the art from this detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

The following drawings form part of the present specification and are included to further demonstrate certain aspects of the present invention. The invention may be better understood by reference to one or more of these drawings in combination with the detailed description of specific embodiments presented herein:

FIG. 1A, FIG. 1B, FIG. 1C and FIG. 1D. Emission of phenylpropanoids from *C. breweri* flowers as measured by headspace collection at 12-h intervals and GC-MS analysis. (FIG. 1A) Emission of eugenol. (FIG. 1B) Emission of isoeugenol (FIG. 1C) Emission of methyleugenol. (FIG. 1D) Emission of methylisoeugenol.

FIG. 2. Emission of phenylpropanoids from *C. breweri* flowers and flower parts as measured by headspace collection at 12-h intervals and GC-MS analysis.

FIG. 3. The reactions catalyzed by IEMT, and the possible pathways leading to eugenol and isoeugenol (based on Manitto *et al.*, 1974; Senanayake *et al.*, 1977).

FIG. 4A, FIG. 4B, and FIG. 4C. Levels of different OMT activities in different parts of the flower during the lifespan of the flower. (FIG. 4A) Eugenol OMT activity. (FIG. 4B) Isoeugenol OMT activity. (FIG. 4C) Caffeic acid OMT activity. Data are shown only for flower parts that contained detectable IEMT activity (fkat = femtomoles of product per second).

FIG. 5. Steps in the purification of OMTs from *C. breweri* petals.

FIG. 6. Nucleotide sequence of IEMT1 cDNA clone. The predicted protein sequence is shown below the nucleotide sequence. Numbers on right refer to the nucleotide sequence, and numbers on left refer to the protein sequence. Peptide sequences obtained experimentally are underlined. The three conserved motifs hypothesized to bind SAM (Kagan and Clarke, 1994) are shown in black boxes with white letters.

FIG. 7A, FIG. 7B and FIG. 7C. (FIG. 7A) Enzymatic activity of the plant purified OMT preparation. Several related substrates in addition to eugenol and isoeugenol were tested as substrates (*see* text and FIG. 3). (FIG. 7B) Activity of plant IEMT1 expressed in *E. coli*. Results are presented as percentage of activity relative to activity with eugenol, which is arbitrarily set at 100%. (FIG. 7C) Activity of plant COMT expressed in *E. coli*. Results are presented as percentage of activity relative to activity with caffeic acid, which is arbitrarily set at 100%.

FIG. 8A and FIG. 8B. Expression of IEMT in flower parts. (FIG. 8A) Northern blot hybridization with mRNA from different tissues of an inbred line of plants that emit methylisoeugenol, using IEMT probe derived from the coding region of IEMT1. (FIG. 8B) Northern blot hybridization with mRNA from different tissues of an inbred line of plants that do not emit methylisoeugenol, using the same IEMT probe as in A. Lanes were loaded with 7 μg of total RNA. Autoradiography was for 48 h. Control lane in B contained 7 μg of total

RNA from petals of methylisoeugenol-emitting flowers (*i.e.*, sample identical to the petal RNA lane in A). Each blot was rehybridized with an 18S rDNA probe to standardize samples.

FIG. 9A and FIG. 9B. Expression of IEMT in petals during flower development. (FIG. 9A) Representative Northern blot hybridization study with mRNAs extracted from petal tissue at different stages in the lifespan of methylisoeugenol-emitting flowers. Each lane contained 3 μg of total RNA. Autoradiography was for 14 days. The blot was rehybridized with an 18S rDNA probe to standardize samples. (FIG. 9B) Plot of the variation in levels of petal IEMT mRNA in methylisoeugenol-emitting flowers over time. Values were obtained by scanning blots with a phosphoimager. Each point is the average of four different studies (including the one shown in A), and values were corrected by standardizing for 18S RNA levels.

FIG. 10. The biochemical reactions catalyzed by IEMT and COMT. IEMT methylates the 4′ hydroxyl group of isoeugenol and eugenol to make isomethyleugenol and methyleugenol, respectively. COMT methylates the 3′ hydroxyl group of caffeic acid and the 5′ hydroxy group of 5′-hydroxylferulic acid to make ferulic acid and sinapic acid, respectively.

FIG. 11. SDS-PAGE of purified IEMT preparations. Lane 1, IEMT co-purified with COMT from plant tissue. Lane 2, IEMT purified from *E. coli*. M, marker lane (molecular weights indicated in left). Approximately 0.07 μg of protein was loaded in each of lane 1 and 2. The gel was stained with silver staining.

FIG. 12. Gel filtration chromatography of IEMT. Purified IEMT from *E. coli* was separated on a gel filtration column QK-PAK TSK GFC 300 GL (Tosohaas, Montgomeryville, PA) and the protein peak was detected by activity assay. Protein standards were Cytochrome C (12.4 k/D), Ovalbumin (45 kD), Bovine Serum Albumin (67 kD), Alcohol Dehydrogenase (150 kD), IgG (169 kD), β-amylase (200 kD) and Apoferritin (443 kDa). Solid line and circles represent protein standard curve. Broken line and triangles represent the IEMT activity.

FIG. 13. Temperature stability of IEMT activity. IEMT was incubated at temperature ranging from 4°C to 65°C for 30 min and then enzyme activity was assayed. The activity of IEMT at 4°C was set as 100%. Data represent the average of two independent assays.

FIG. 14. pH optimum of IEMT activity. Enzyme assays were carried out with pH values raging from 6.0 to 10.0 in two buffer systems with overlapped pH ranges. Dots represent Bis-Tris buffer system and circles represent Tris buffer system. IEMT activity assayed in Tris-HC1 buffer pH 7.5 was set as 100%. Data represent the average of three independent assays.

FIG. 15. Amino acid sequence alignment of IEMT and COMT from *C. breweri,* and COMT from aspen (SEQ ID NO:11), and accession numbers AF006009 (SEQ ID NO:12) and X62096 (SEQ ID NO:13), respectively). Numbers on right refer to the protein sequence. Dots represent residues identical to the top sequence (IEMT). Numbers delineating the beginning and the end of substituted segments are indicated. The three conserved SAM binding motifs are boxed.

FIG. 16. Schematic representation of different chimerical protein constructs and their kinetic parameters. Data for aspen COMT are calculated from Meng and Campbell (1996).

FIG. 17. *In vitro* mutagenesis of IEMT. Some amino acids when changed, reduce IEMT activity, showing their importance in IEMT catalyctic acitivity and/or substrate specificity.

DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

In an effort to provide the means for producing plants with altered flavor and fragrance characteristics, the current inventors have sought to exploit and manipulate the biosynthetic

pathway of phenylpropanoids. In one aspect, the invention relates to the cloning of a gene responsible for the production phenylpropanoid components of the *C. breweri* scent. These findings provide for methodology permitting production of transgenic plants with altered fragrance and flavoring characteristics.

Here the inventors report that petal tissue was found to be responsible for the bulk of the phenylpropanoid emission. The activity of *S*-adenosyl-L-methionine:(iso)eugenol *O*-methyltransferase (IEMT), a novel enzyme that catalyzes the methylation of the *para*-4'-hydroxyl of both eugenol and isoeugenol to methyleugenol and methylisoeugenol, respectively, also was highest in petal tissue. IEMT activity was absent from floral tissues of plants not emitting methyl(iso)eugenol. A *C. breweri* cDNA clone encoding IEMT was isolated, and its sequence was shown to have 70% identity to *S*-adenosyl-L-methionine:caffeic acid *O*-methyltransferase (COMT). The protein encoded by the IEMT cDNA can utilize eugenol and isoeugenol as substrates, but not caffeic acid or various of the other phenylpropanoid pathway components. Steadystate IEMT mRNA levels were positively correlated with levels of IEMT activity in the tissues, and no IEMT mRNA was observed in flowers that do not emit methyl(iso)eugenol. Overall, these data show that the floral emission of methyl(iso)eugenol is controlled at the site of emission, that a positive correlation exists between volatile emission and IEMT activity, and that control of the level of IEMT activity is exerted at a pretranslational step.

IEMT is the first enzyme to have been identified that possesses the capability of methylating compounds of the lignin biosynthetic pathway at position C4 (see FIG. 3). This enzyme possesses high sequence similarity to COMTs, but it cannot utilize caffeic acid as a substrate. Conversely, COMT can use caffeic acid, but it has a 3-methyl transferase activity and cannot utilize (iso)eugenol as a substrate. However, using the methods and compositions described herein, it will be possible to engineer such 4-methyl transferase activity into COMT and caffeic acid specificty into IEMT. For example, amino acid residues 92-173 of COMT appear to be responsible for the substrate specificity of this molecule and also contain the -OH

group upon which the donor methyl group is carried prior to its transfer to the substrate. Thus, manipulations involving this region in particular will prove useful.

The present invention, in one embodiment, provides an isolated IEMT polypeptide (SEQ ID NO:2). The present invention further provides an isolated nucleic acid encoding an IEMT polypeptide. This isolated nucleic acid may be derived from *C. breweri* or various monocot or dicot plants. Disclosed here are methods and compositions for using the IEMT polypeptide and nucleotides that encode IEMT. These methods and compositions may involve expression of a DNA sequence encoding an IEMT polypeptide linked operably to a promoter functional in a plant cell. The use of such methods and compositions result in the alteration in a particular fragrance quality or flavor of a plant. In particular, increase in IEMT activity is associated with a decrease in eugenol and isoeugenol, which are responsible for the clove-like flavor of some plants.

With the identification of the IEMT gene as key gene in the phenylpropanoid pathway, a number of different endeavors become possible. For example, one may employ both protein and DNA compositions in the production of transgenic plants with improved characteristics and greater viablity. More specifically, it will be possible to engineer plants to exhibit a reduced clove-like taste by promoting the methylation of eugenol and isoeugenol into methyleugenol and methylisoeugenol, respectively. It also will be possible to determine which domain(s) of IEMT is(are) responsible for methyl transfer activity. This will involve the production of truncated, deletion, fusion and replacement mutants of IEMT. A candidate parter for fusions and replacements is COMT. In this way, new enzymes with unique activities and/or specificities will be produced.

In particular, methylation of the 4'-para-hydroxyl group in lignin precursors such as caffeic acid, ferulic acid, coniferyl aldehyde and coniferyl alcohol cannot be accomplished using caffeic acid methyl transferase (COMT) because of steric hindrance from the functional groups attached to C7 position of these compounds. Using the 4-para-hydroxy methylation action of IEMT disclosed herein, one can produce a fusion proteins or chimeras of, for example,

COMT and IEMT, to produce a polypeptide capable of methylating caffeic acid and its aforementioned derivatives at the 4-hydroxyl group, resulting in a block in lignin production.

In yet other examples, it is possible that the products and by-products of the phenylpropanoid pathway are involved in conferring insect attractant properties, thereby facilitating pollenation in important crop plants that require insect pollination. The methods and compositions for achieving these and other objectives are detailed herein below.

I. Biosynthesis of Phenylpropanoids and Lignins

Lignins are the product of a dehydrogenative polymerization of three primary precursors - the trans-coniferyl, trans-sinapyl and trans-p-coumaryl alcohols. The monomers can occur in lignins in different proportions and with different types of linkages both with each other and with the surrounding cell wall polysaccharides, thus producing a wide variety of polymers. These polymers, or "lignin cores," always are associated covalently with hemicelluloses. Most lignins also contain varying amounts of aromatic carboxylic acids in ester-like combinations. Such differences in the structure of lignins are usually found in plant species. However, differences in the composition of lignins, and even in the binding to the primary and secondary cell walls, also can occur in the same plant between different tissues of different ages. The biosynthesis of lignin monomers is a part of the phenylpropanoid biosynthesis pathway, which also is responsible for the production of a wide range of compounds including flavonoid pigments, isoflavonoids, coumarin phytoalexins and cell division promoting dehydroiconiferyl glucosides.

Phenylalanine is deaminated to produce cinnamic acid. This acid is then hydroxylated and methylated, producing different acids substituted on the aromatic ring. Coenzyme A thioesters of (p)-coumaric, ferulic and sinapic acids then are produced by the action of hydroxycinnamate:CoA ligase. These compounds are subsequently reduced by cinnamyl-CoA reductase (CCR) to cinnamaldehydes, which are finally converted to cinnamyl alcohols by the cinnamyl alcohol dehydrogenase (CAD). Only the last two reactions are specific for the

biosynthesis of lignin. The cinnamyl alcohols are then believed to be transported to the cell wall where they are polymerized by peroxidase in the presence of hydrogen peroxide.

When the surface growth of the cell ceases, it is followed by a phase of wall thickening (secondary wall formation). Lignification takes place predominantly during this phase. It starts in the cell corners and extends along the middle lamella, through the primary wall and, finally, to the secondary wall. External factors can induce qualititave and quantitative modification in lignification. The synthesis of new types of lignins, sometimes in tissues which are not normally lignified, may be induced by infection with pathogenic microorganisms. Lignification is stimulated by light, as well as by low calcium levels, by boron, by mechanical stress and by infection.

Cinnamyl alcohol dehydrogenase: (CAD, E.C. 1.1.1.195) catalyses of conversion of cinnamaldehydes to cinnamyl alcohols. CAD has been characterized for several different species: *Forsythia suspensa*, soybean (*Glycine max*), spruce (*Picea abies*), poplar (*Populus euramericana*) and eucalyptus and reviewed in U.S. Patent No. 5,451,514, specifically incorporated herein by reference.

An alternate fate of coumaric acid, rather than being converted to lignin, is to be hydroxylated at the C3 position, which yields caffeic acid. This product then may be methylated at the C3 position by COMT and enter the pathway of phenylpropanoid biosynthesis, as depicted in FIG. 3. However, these compounds are not completely committed to phenylpropanoid biosynthesis and may still feed into the lignin biosynthetic pathway. The product of COMT is ferulic acid, which is converted to coniferyl aldehyde and coniferyl alcohol.

Based on this pathway, it is possible to decrease lignification by increasing the amount of precursors that are available for methyl(iso)eugenol biosynthesis. Thus, by altering IEMT activity to accept additional substrates, it is possible to decrease lignification, as lignin precursors that are methylated at the 4' -OH group cannot be polymerized to form lignin.

II. IEMT Polypeptides and Fragments Thereof

Thus, according to one aspect of the present invention, the present inventors provide a novel enzyme, IEMT. This molecule will prove useful in a variety of different contexts. For example, IEMT may be used in catalyzing the methyl transfer reaction in test samples. IEMT also can be used as part of a screening assay to examine reagents for their ability to affect methyl transferation *in vitro*. In another embodiment, IEMT will prove useful in standard laboratory procedures as protein marker of known molecular weight.

In addition to the entire IEMT molecule, the present invention also relates to fragments of the polypeptide that may or may not retain the methyl transferase (or other) activity. Fragments including the N-terminus of the molecule may be generated by genetic engineering of translation stop sites within the coding region (discussed below). Alternatively, treatment of the IEMT molecule with proteolytic enzymes, known as protease, can produces a variety of N-terminal, C-terminal and internal fragments. Examples of fragments may include contiguous residues of the IEMT sequence given in SEQ ID NO:2 of 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 55, 60, 65, 75, 80, 85, 90, 95, 100, 200, 300, 400 or more amino acids in length. These fragments may be purified according to known methods, such as precipitation (*e.g.*, ammonium sulfate), HPLC, ion exchange chromatography, affinity chromatography (including immunoaffinity chromatography) or various size separations (sedimentation, gel electrophoresis, gel filtration).

A. Structural Features of the Polypeptide

The inventors have purified IEMT, a protein with an apparent molecular mass of 40 kD. On separation and closer inspection, using N-terminal sequence analysis, it was found that there were two proteins having very similar peptide sequences (from the top protein band: XTGNAETQLTP [SEQ. ID NO:3, X = unidentified]; from the bottom protein band: SPGNAEIQIIP [SEQ. ID NO:4]), both of which show some similarity to dicotyledonous plant COMT N-terminal sequences. SEQ ID NO:9 sets forth an exemplary protein sequence of

COMT encoded by the COMT gene set forth in SEQ ID NO:8. Other exemplary gene sequences for COMT can be found on Genbank, for example Genbank accession numbers U16793, X62096 and M72523.

Additionally, several individual peptides were obtained from SDS-PAGE after cyanogen bromide cleavage of a mixture of the two proteins. These sequences also showed significant similarity to COMT sequences. None of the peptide sequences showed any similarity to CCOMT, an enzyme that methylates caffeic acid bound to CoA in an alternative lignin biosynthesis pathway. Interestingly, this enzyme bears no significant similarity to COMT (Ye *et al.*, 1994).

Since the N-terminal sequences of both proteins in the "purified" IEMT preparation showed similarity to COMT, the inventors used a COMT cDNA clone from aspen (Bugos *et al.*, 1991) as a probe in low-stringency hybridization screening of a *C. breweri* flower cDNA library. Several clones were isolated, and the nucleotide sequence of one of them, designated IEMT1, was determined and is shown in SEQ ID NO:1 and in FIG. 6. This clone contains 1486 nucleotides, not including the poly(A) tail, with an open reading frame of 368 codons, beginning with an ATG codon at positions 43 to 45. There is one stop codon in-frame upstream of this ATG, indicating that this cDNA clone contains the entire coding region. This conclusion further is supported by primer-extension studies that determined the 5′ end of the mRNA. The molecular mass of the protein encoded by the open reading frame of IEMT1 is 40 kD, the same as that of the two proteins found in the purified OMT preparation.

In addition, the protein encoded by IEMT1 contains two of the internal peptide sequences determined experimentally, MLDRVLRLLASYSVVTYTLRE (SEQ. ID NO:5) and MFDGVPKGDAIFIK (SEQ. ID NO:6). The IEMT1 protein is approximately 30% divergent from all available dicot COMT sequences. These COMTs vary among themselves by no more than 15%, with the exception of a *Zinnia elegans* sequence, designated COMT, that has a substrate specificity that has not been extensively tested (Ye and Varner, 1995). IEMT1 shows sequence identity of 65% or less to several other types of plant OMTs in the data bank, and it

contains the three conserved motifs (FIG. 6) identified by Kagan and Clarke (1994) and hypothesized by these authors to be involved in the binding of SAM.

B. Functional Characteristics

The inventors previously have shown that the strong, sweet fragrance of *Clarkia breweri* (Onagraceae) consists of 8 to 12 volatile compounds, including four phenylpropanoids. Although some *C. breweri* plants emit all four phenylpropanoids (eugenol, isoeugenol, methyleugenol and methylisoeugenol), other *C. breweri* plants do not emit the latter two compounds. Little is known about the enzymatic steps leading to the synthesis of eugenol and isoeugenol, although pulse-chase studies with radioactive precursors have clearly shown that they are ultimately derived from phenylalanine, with *p*-coumaric acid as an intermediate (Manitto *et al.*, 1974; Senanayake *et al.*, 1977).

The inventors' investigation concentrated on the biosynthesis of methyl(iso)eugenol *via* the action of IEMT. To date, no enzymatic activity capable of converting eugenol and isoeugenol to methyleugenol and methylisoeugenol, respectively, has been reported from plants. The present invention demonstrates the existence of a single enzyme, designated IEMT, which methylates both eugenol and isoeugenol to methyleugenol and methylisoeugenol, respectively, with high specificity. Although this enzyme has high sequence similarity to COMTs, it cannot use caffeic acid as a substrate, nor can COMT use (iso)eugenol as a substrate.

As stated above, levels of IEMT activity and mRNA in the different floral tissues of methyl(iso)eugenol emitters strongly correlate with the production and emission of these two compounds by the same tissues, being highest in petals, followed by stamens, style, and stigma, and absent in sepals and in leaf and stem tissue. Moreover, non-emitting plants did not have IEMT activity or IEMT mRNA in any floral tissues, although they do contain the IEMT gene in their genome. These results are very similar to those obtained for LIS, an enzyme that produces linalool, another floral scent compound in *C. breweri*. The inventors observed strong positive

correlation between levels of LIS enzyme activity, protein, and mRNA at the site of synthesis and emission of linalool in the flower (Pichersky *et al.*, 1994; Dudareva *et al.*, 1996).

The inventors have shown that methylisoeugenol and methyleugenol emission, IEMT activity, and mRNA levels in the petals all increase in parallel as buds mature and flowers open (again with mRNA levels peaking 1-2 days ahead of enzyme activity and emission). However, starting from the third day of anthesis (one day after the stigma becomes receptive and most pollination occurs), emission declines but IEMT activity remains relatively stable, and IEMT mRNA levels actually increased after declining 25% from their peak on the day before anthesis.

The consequence of high levels of IEMT activity without methyl(iso)eugenol emission in older flowers is unclear. It could be that methyleugenol and methylisoeugenol still are being made by the floral tissues but are being tied into non-volatile compounds. Loughrin *et al.* (1992) have reported an increase in glycosidically bound scent compounds in tobacco floral tissues as the flowers aged. Although methyleugenol and methylisoeugenol cannot be linked directly to a sugar moiety, they may be conjugated to other compounds. Alternatively, methylisoeugenol and methyleugenol biosynthesis may decline as the flower ages, even though IEMT levels remain high, because of changes in other factors involved (*e.g.*, a decline in levels of other enzymes in the pathway, especially rate-limiting enzymes).

When the present application refers to the function of IEMT or "wild-type" activity, it is meant that the enzyme has the ability to mediate the methyl transfer reaction at the C4-position of (iso)eugenol to produce methyl(iso)eugenol. Thus the molecule in question has the ability to methylate the 4-position of ferulic-coniferyl type molecules of the lignin biosynthetic pathway. Determination of which polypeptides possess this activity may be achieved using assays familiar to those of skill in the art.

For example, cell extracts may be assayed for by transfecting cells with the test cDNA in an appropriate vector and then assaying these cells for methyl transferase activity. Assays of

methyl transferase activity are well known to those of skill in the art and are described in, for example, by DeCarolis and Ibrahim (1989) and in Example 1.

C. Variants of IEMT

The present invention contemplates variants of IEMT. Amino acid sequence variants of the polypeptide can be substitutional, insertional or deletion variants. Deletion variants lack one or more residues of the native protein which are not essential for function or immunogenic activity, and are exemplified by the variants lacking a transmembrane sequence described above. Another common type of deletion variant is one lacking secretory signal sequences or signal sequences directing a protein to bind to a particular part of a cell.

Insertional mutants typically involve the addition of material at a non-terminal point in the polypeptide. This may include the insertion of an epitope or simply a single residue. Terminal additions, called fusion proteins, are discussed below. Insertional mutagenesis is based on the inactivation of a gene *via* insertion of a known DNA fragment. Because it involves the insertion of some type of DNA fragment, the mutations generated are generally loss-of-function rather than gain-of-function mutations. However, there are several examples of insertions generating gain-of-function mutations (Oppenheimer *et al.* 1991). Insertion mutagenesis has been very successful in bacteria and *Drosophila* (Cooley *et al.* 1988) and recently has become a powerful tool in several plant species (corn; *e.g.*, Schmidt *et al.* 1987); *Arabidopsis*; *e.g.*, Marks *et al.*, 1991; Koncz *et al.* 1990; *Antirrhinum*: *e.g.*, Sommer *et al.* 1990).

Substitutional variants typically contain the exchange of one amino acid for another at one or more sites within the protein, and may be designed to modulate one or more properties of the polypeptide, such as stability against proteolytic cleavage, without the loss of other functions or properties. Substitutions of this kind preferably are conservative, that is, one amino acid is replaced with one of similar shape and charge. Conservative substitutions are well known in the art and include, for example, the changes of: alanine to serine; arginine to lysine; asparagine to glutamine or histidine; aspartate to glutamate; cysteine to serine; glutamine to asparagine; glutamate to aspartate; glycine to proline; histidine to asparagine or glutamine; isoleucine to

leucine or valine; leucine to valine or isoleucine; lysine to arginine; methionine to leucine or isoleucine; phenylalanine to tyrosine, leucine or methionine; serine to threonine; threonine to serine; tryptophan to tyrosine; tyrosine to tryptophan or phenylalanine; and valine to isoleucine or leucine.

The following is a discussion based upon changing of the amino acids of a protein to create an equivalent, or even an improved, second-generation molecule. For example, certain amino acids may be substituted for other amino acids in a protein structure without appreciable loss of interactive binding capacity with structures such as, for example, antigen-binding regions of antibodies or binding sites on substrate molecules. Since it is the interactive capacity and nature of a protein that defines that protein's biological functional activity, certain amino acid substitutions can be made in a protein sequence, and its underlying DNA coding sequence, and nevertheless obtain a protein with like properties. It is thus contemplated by the inventors that various changes may be made in the DNA sequences of genes without appreciable loss of their biological utility or activity, as discussed below. Table 1 shows the codons that encode particular amino acids.

In making such changes, the hydropathic index of amino acids may be considered. The importance of the hydropathic amino acid index in conferring interactive biologic function on a protein is generally understood in the art (Kyte and Doolittle, 1982). It is accepted that the relative hydropathic character of the amino acid contributes to the secondary structure of the resultant protein, which in turn defines the interaction of the protein with other molecules, for example, enzymes, substrates, receptors, DNA, antibodies, antigens, and the like.

Each amino acid has been assigned a hydropathic index on the basis of their hydrophobicity and charge characteristics (Kyte and Doolittle, 1982), these are: isoleucine (+4.5); valine (+4.2); leucine (+3.8); phenylalanine (+2.8); cysteine/cystine (+2.5); methionine (+1.9); alanine (+1.8); glycine (-0.4); threonine (-0.7); serine (-0.8); tryptophan (-0.9); tyrosine (-1.3); proline (-1.6); histidine (-3.2); glutamate (-3.5); glutamine (-3.5); aspartate (-3.5); asparagine (-3.5); lysine (-3.9); and arginine (-4.5).

It is known in the art that certain amino acids may be substituted by other amino acids having a similar hydropathic index or score and still result in a protein with similar biological activity, *i.e.*, still obtain a biological functionally equivalent protein. In making such changes, the substitution of amino acids whose hydropathic indices are within ±2 is preferred, those which are within ±1 are particularly preferred, and those within ±0.5 are even more particularly preferred.

It is also understood in the art that the substitution of like amino acids can be made effectively on the basis of hydrophilicity. U.S. Patent 4,554,101, incorporated herein by reference, states that the greatest local average hydrophilicity of a protein, as governed by the hydrophilicity of its adjacent amino acids, correlates with a biological property of the protein. As detailed in U.S. Patent 4,554,101, the following hydrophilicity values have been assigned to amino acid residues: arginine (+3.0); lysine (+3.0); aspartate (+3.0 ± 1); glutamate (+3.0 ± 1); serine (+0.3); asparagine (+0.2); glutamine (+0.2); glycine (0); threonine (-0.4); proline (-0.5 ± 1); alanine (-0.5); histidine *-0.5); cysteine (-1.0); methionine (-1.3); valine (-1.5); leucine (-1.8); isoleucine (-1.8); tyrosine (-2.3); phenylalanine (-2.5); tryptophan (-3.4).

It is understood that an amino acid can be substituted for another having a similar hydrophilicity value and still obtain a biologically equivalent and immunologically equivalent protein. In such changes, the substitution of amino acids whose hydrophilicity values are within ±2 is preferred, those that are within ±1 are particularly preferred, and those within ±0.5 are even more particularly preferred.

As outlined above, amino acid substitutions are generally based on the relative similarity of the amino acid side-chain substituents, for example, their hydrophobicity, hydrophilicity, charge, size, and the like. Exemplary substitutions that take various of the foregoing characteristics into consideration are well known to those of skill in the art and include: arginine and lysine; glutamate and aspartate; serine and threonine; glutamine and asparagine; and valine, leucine and isoleucine.

Another embodiment for the preparation of polypeptides according to the invention is the use of peptide mimetic. Mimetic are peptide-containing molecules that mimic elements of protein secondary structure. See, for example, Johnson *et al.*, "Peptide Turn Mimetic" in *BIOTECHNOLOGY AND PHARMACY*, Pezzuto *et al.*, Eds., Chapman and Hall, New York (1993). The underlying rationale behind the use of peptide mimetic is that the peptide backbone of proteins exists chiefly to orient amino acid side chains in such a way as to facilitate molecular interactions, such as those of antibody and antigen. A peptide mimetic is expected to permit molecular interactions similar to the natural molecule. These principles may be used, in conjunction with the principles outline above, to engineer second generation molecules having many of the natural properties of IEMT, but with altered and even improved characteristics.

D. Domain Switching

As described in the Examples, the present inventors have identified a novel methyltransferase which catalyzes the addition of a methyl group to the C4 position of (iso)eugenol to yield methyl(iso)eugenol. There is a reasonable expectation that other homologs, allelic variants and mutants of this gene exist in related species, such as basil, clove, tomato, potato, corn, soybean, oil rapeseed, wheat, barley, rye and others. Upon isolation of these homologs, variants and mutants, and in conjunction with other analyses, certain active or functional domains can be identified. This will provide a starting point for further mutational analysis of the molecule. One way in which this information can be exploited is in "domain switching."

Domain switching involves the generation of chimeric molecules using different but, in this case, related polypeptides. By comparing the sequences for IEMT with the other methyl transferases, for example COMT, and possibly with mutants and allelic variants of these polypeptides, one can make predictions as to the functionally significant regions of these molecules. It then is possible, to switch related domains of these molecules in an effort to determine the criticality of these regions to IEMT function. These molecules may have additional value in that the "chimeras" can be distinguished from natural molecules, while

possibly providing the same function. In a particular embodiment, this aspect of the invention will be useful in genetically engineering for example, COMT with portions of IEMT, and *vice versa*, in order to achieve methylation of molecules not normally found to be natural substrates of one or both enzymes. For example, IEMT could be engineered to C4-methylate caffeic acid, or any other ferulic acid, coniferyl alcohol or coniferyl aldehyde in the lignin biosynthetic pathway. Similarly, COMT can be modified to methylate these molecules at the C4 position.

Based on the sequence identity, at the amino acid level, it may be inferred that even small changes in the primary sequence of the molecule will affect function. Further analysis of mutations and their predicted effect on secondary structure will add to this understanding. In one such study, it has been determined that residues 92-173 of COMT confers substrate specificity on this enzyme, and also contains a key hydroxyl moiety from which the methyl group is donated to the substrate.

It is proposed, therefore, that a series of chimera will be generated. Primarily, the convenience of restriction sites will dictate the precise location of breakpoints. However, generally, one may produce the following types of chimeras. The 5' half of the chimera (residue 1 to about residue 180) may be derived from IEMT, while the 3' half of the molecule may be derived from COMT (about residue 181 to residue 365). Conversely, the 5' half of the chimera (residue 1 to about residue 180) may be derived from COMT, while the 3' half of the molecule may be derived from IEMT (about residue 181 to residue 367). Alternatively, the molecules may be divided roughly into thirds or fourths, having break points at about residues 120 and 240 and 90, 180 and 270, respectively. These chimeras would be arranged in the following manner:

COMT-COMT-IEMT	COMT-IEMT-COMT
IEMT-COMT-COMT	IEMT-IEMT-COMT
IEMT-COMT-IEMT	COMT-IEMT-IEMT
and	
COMT-COMT-COMT-IEMT	COMT-COMT-IEMT-COMT

COMT-IEMT-COMT-IEMT	IEMT-COMT-COMT-COMT
IEMT-IEMT-IEMT-COMT	IEMT-IEMT-COMT-IEMT
IEMT-COMT-IEMT-IEMT	COMT-IEMT-IEMT-IEMT
COMT-IEMT-COMT-IEMT	IEMT-COMT-IEMT-COMT
IEMT-COMT-COMT-IEMT	COMT-IEMT-IEMT-COMT

Another approach is to target particular functional regions of these molecules. For example, the inventors have determined that an 82 amino acid stretch of COMT, when transferred into a corresponding region of IEMT, confers COMT-like activity on the resulting chimera. Thus, it will be of interest to focus on the is region in generating additional chimeras. For example, it may prove desireable to perform random mutagenesis on this region and screen the result clones for the desired activity. It may well be preferable to further define the regions involved by creating more subtle chimeras, for example, in which smaller regions of COMT are transferred into IEMT. Candidate regions from COMT include, approximately, residues 90-100, 100-110, 90-110, 110-120, 90-120, 120-130, 90-130, 110-130, 130-140, 90-140, 140-150, 90-150, 130-150, 150-160, 90-160, 130-160, 160-170, 90-170, 150-170, 170-180, 150-180, 140-180, 130-180, 120-180, 110-180 and 100-180. When placed into the appropriate context on IEMT, these residues may confer upon the chimera the ability to methylate key lignin precursors at the C4 position.

Yet other functionally delineated regions of IEMT are the putative SAM binding sites located at residues 207-215, 260-267 and 290-299. It is likely that these regions should be conserved, at least to a certain extent given that SAM binding is a function that the chimera will need to retain. Nonetheless, these residues may constitute regions of the molecule that, spatially, are in close relation to the substrate molecule and, hence, are involved, directly or indirectly, in substrate specificity.

In vitro mutagenesis studies revealed that TAT at residues 133-135, FL at residues 130-13 and NE at residues 164-165 of SEQ ID NO:10 are important for IEMT catalyctic acitivity and/or substrate specificity, additional mutagensis studies may reveal additional

residues in physical proximity to these residues that are important to the IEMT activity of the molecule. By physical proximity, it is intended that such an addional amino acid may be located close to these amino acids either in the amino acid chain or as part of the tertiary protein structure.

E. Fusion Proteins

A specialized kind of insertional variant is the fusion protein. This molecule generally has all or a substantial portion of the native molecule, linked at the N- or C-terminus, to all or a portion of a second polypeptide. For example, fusions typically employ leader sequences from other species to permit the recombinant expression of a protein in a heterologous host. Another useful fusion includes the addition of a immunologically active domain, such as an antibody epitope, to facilitate purification of the fusion protein. Inclusion of a cleavage site at or near the fusion junction will facilitate removal of the extraneous polypeptide after purification. Other useful fusions include linking of functional domains, such as active sites from enzymes, glycosylation domains, cellular targeting signals or transmembrane regions. Fusion to a polypeptide that can be used for purification of the substrate-IEMT complex would serve to isolated the substrate for identification and analysis.

F. Purification of Proteins

It will be desirable to purify IEMT or variants thereof. Protein purification techniques are well known to those of skill in the art. These techniques involve, at one level, the crude fractionation of the cellular milieu to polypeptide and non-polypeptide fractions. Having separated the polypeptide from other proteins, the polypeptide of interest may be further purified using chromatographic and electrophoretic techniques to achieve partial or complete purification (or purification to homogeneity). Analytical methods particularly suited to the preparation of a pure peptide are ion-exchange chromatography, exclusion chromatography; polyacrylamide gel electrophoresis; isoelectric focusing. A particularly efficient method of purifying peptides is fast protein liquid chromatography or even HPLC.

Certain aspects of the present invention concern the purification, and in particular embodiments, the substantial purification, of an encoded protein or peptide. The term "purified protein or peptide" as used herein, is intended to refer to a composition, isolatable from other components, wherein the protein or peptide is purified to any degree relative to its naturally-obtainable state. A purified protein or peptide therefore also refers to a protein or peptide, free from the environment in which it may naturally occur.

Generally, "purified" will refer to a protein or peptide composition that has been subjected to fractionation to remove various other components, and which composition substantially retains its expressed biological activity. Where the term "substantially purified" is used, this designation will refer to a composition in which the protein or peptide forms the major component of the composition, such as constituting about 50%, about 60%, about 70%, about 80%, about 90%, about 95% or more of the proteins in the composition.

Various methods for quantifying the degree of purification of the protein or peptide will be known to those of skill in the art in light of the present disclosure. These include, for example, determining the specific activity of an active fraction, or assessing the amount of polypeptides within a fraction by SDS/PAGE analysis. A preferred method for assessing the purity of a fraction is to calculate the specific activity of the fraction, to compare it to the specific activity of the initial extract, and to thus calculate the degree of purity, herein assessed by a "-fold purification number." The actual units used to represent the amount of activity will, of course, be dependent upon the particular assay technique chosen to follow the purification and whether or not the expressed protein or peptide exhibits a detectable activity.

Various techniques suitable for use in protein purification will be well known to those of skill in the art. These include, for example, precipitation with ammonium sulphate, PEG, antibodies and the like or by heat denaturation, followed by centrifugation; chromatography steps such as ion exchange, gel filtration, reverse phase, hydroxylapatite and affinity chromatography; isoelectric focusing; gel electrophoresis; and combinations of such and other techniques. As is generally known in the art, it is believed that the order of conducting the

various purification steps may be changed, or that certain steps may be omitted, and still result in a suitable method for the preparation of a substantially purified protein or peptide.

There is no general requirement that the protein or peptide always be provided in their most purified state. Indeed, it is contemplated that less substantially purified products will have utility in certain embodiments. Partial purification may be accomplished by using fewer purification steps in combination, or by utilizing different forms of the same general purification scheme. For example, it is appreciated that a cation-exchange column chromatography performed utilizing an HPLC apparatus will generally result in a greater "-fold" purification than the same technique utilizing a low pressure chromatography system. Methods exhibiting a lower degree of relative purification may have advantages in total recovery of protein product, or in maintaining the activity of an expressed protein.

It is known that the migration of a polypeptide can vary, sometimes significantly, with different conditions of SDS/PAGE (Capaldi *et al.*, 1977). It will therefore be appreciated that under differing electrophoresis conditions, the apparent molecular weights of purified or partially purified expression products may vary.

High Performance Liquid Chromatography (HPLC) is characterized by a very rapid separation with extraordinary resolution of peaks. This is achieved by the use of very fine particles and high pressure to maintain an adequate flow rate. Separation can be accomplished in a matter of minutes, or at most an hour. Moreover, only a very small volume of the sample is needed because the particles are so small and close-packed that the void volume is a very small fraction of the bed volume. Also, the concentration of the sample need not be very great because the bands are so narrow that there is very little dilution of the sample.

Gel chromatography, or molecular sieve chromatography, is a special type of partition chromatography that is based on molecular size. The theory behind gel chromatography is that the column, which is prepared with tiny particles of an inert substance that contain small pores, separates larger molecules from smaller molecules as they pass through or around the pores,

depending on their size. As long as the material of which the particles are made does not adsorb the molecules, the sole factor determining rate of flow is the size. Hence, molecules are eluted from the column in decreasing size, so long as the shape is relatively constant. Gel chromatography is unsurpassed for separating molecules of different size because separation is independent of all other factors such as pH, ionic strength, temperature, *etc.* There also is virtually no adsorption, less zone spreading and the elution volume is related in a **simple** matter to molecular weight.

Affinity Chromatography is a chromatographic procedure that relies on the specific affinity between a substance to be isolated and a molecule that it can specifically bind to. This is a receptor-ligand type interaction. The column material is synthesized by covalently coupling one of the binding partners to an insoluble matrix. The column material is then able to specifically adsorb the substance from the solution. Elution occurs by changing the conditions to those in which binding will not occur (alter pH, ionic strength, temperature, *etc.*).

A particular type of affinity chromatography useful in the purification of carbohydrate containing compounds is lectin affinity chromatography. Lectins are a class of substances that bind to a variety of polysaccharides and glycoproteins. Lectins are usually coupled to agarose by cyanogen bromide. Conconavalin A coupled to Sepharose was the first material of this sort to be used and has been widely used in the isolation of polysaccharides and glycoproteins other lectins that have been include lentil lectin, wheat germ agglutinin which has been useful in the purification of N-acetyl glucosaminyl residues and *Helix pomatia* lectin. Lectins themselves are purified using affinity chromatography with carbohydrate ligands. Lactose has been used to purify lectins from castor bean and peanuts; maltose has been useful in extracting lectins from lentils and jack bean; N-acetyl-D galactosamine is used for purifying lectins from soybean; N-acetyl glucosaminyl binds to lectins from wheat germ; D-galactosamine has been used in obtaining lectins from clams and L-fucose will bind to lectins from lotus.

The matrix should be a substance that itself does not adsorb molecules to any significant extent and that has a broad range of chemical, physical and thermal stability. The ligand should

be coupled in such a way as to not affect its binding properties. The ligand should also provide relatively tight binding. And it should be possible to elute the substance without destroying the sample or the ligand. One of the most common forms of affinity chromatography is immunoaffinity chromatography. The generation of antibodies that would be suitable for use in accord with the present invention is discussed below.

G. Synthetic Peptides

The present invention also describes smaller IEMT-related peptides for use in various embodiments of the present invention. Because of their relatively small size, the peptides of the invention can also be synthesized in solution or on a solid support in accordance with conventional techniques. Various automatic synthesizers are commercially available and can be used in accordance with known protocols. See, for example, Stewart and Young, (1984); Tam *et al.*, (1983); Merrifield, (1986); and Barany and Merrifield (1979), each incorporated herein by reference. Short peptide sequences, or libraries of overlapping peptides, usually from about 6 up to about 35 to 50 amino acids, which correspond to the selected regions described herein, can be readily synthesized and then screened in screening assays designed to identify reactive peptides. Alternatively, recombinant DNA technology may be employed wherein a nucleotide sequence which encodes a peptide of the invention is inserted into an expression vector, transformed or transfected into an appropriate host cell and cultivated under conditions suitable for expression.

III. Nucleic Acids

The present invention also provides, in another embodiment, DNAs encoding IEMT or fragments thereof. The present invention identifies the gene for the *C. breweri* IEMT molecule, however, the present invention is not limited in scope to this gene, as one of ordinary skill in the could, using the nucleic acids disclosed herein, readily identify related homologs in various other species (*e.g.*, basil, clove, tomato, potato, corn, soybean and others).

In addition, it should be clear that the present invention is not limited to the specific nucleic acids disclosed herein. As discussed below, an "IEMT gene" may contain a variety of different bases and yet still produce a corresponding polypeptides that is functionally indistinguishable, and in some cases structurally, from the *C. breweri* gene disclosed herein.

Similarly, any reference to a nucleic acid should be read as encompassing a host cell containing that nucleic acid and, in some cases, capable of expressing the product of that nucleic acid. Cells expressing nucleic acids of the present invention may prove useful in the context of screening for agents that induce, repress, inhibit, augment, interfere with, block, abrogate, stimulate or enhance the function of IEMT.

A. Nucleic Acids Encoding IEMT

The gene disclosed in SEQ ID NO: 1 is the IEMT1 gene of the present invention. Nucleic acids according to the present invention may encode an entire IEMT gene, a domain of IEMT that expresses function capable of mediating methyltransferase activity, or any other fragment of the IEMT sequences set forth herein. The nucleic acid may be derived from genomic DNA, *i.e.*, cloned directly from the genome of a particular organism. In preferred embodiments, however, the nucleic acid would comprise complementary DNA (cDNA). Also contemplated is a cDNA plus a natural intron or an intron derived from another gene; such engineered molecules are sometime referred to as "mini-genes." At a minimum, these and other nucleic acids of the present invention may be used as molecular weight standards in, for example, gel electrophoresis.

The term "cDNA" is intended to refer to DNA prepared using messenger RNA (mRNA) as template. The advantage of using a cDNA, as opposed to genomic DNA or DNA polymerized from a genomic, non- or partially-processed RNA template, is that the cDNA primarily contains coding sequences of the corresponding protein. There may be times when the full or partial genomic sequence is preferred, such as where the non-coding regions are required for optimal expression or where non-coding regions such as introns are to be targeted in an antisense strategy.

It also is contemplated that a given IEMT from a given species may be represented by natural variants that have slightly different nucleic acid sequences but, nonetheless, encode the same protein (see Table 1 below).

As used in this application, the term "a nucleic acid encoding a IEMT" refers to a nucleic acid molecule that has been isolated free of total cellular nucleic acid. In preferred embodiments, the invention concerns a nucleic acid sequence essentially as set forth in SEQ ID NO:1. The term "as set forth SEQ ID NO:1" means that the nucleic acid sequence substantially corresponds to a portion of SEQ ID NO:1. The term "functionally equivalent codon" is used herein to refer to codons that encode the same amino acid, such as the six codons for arginine or serine (Table 1, below), and also refers to codons that encode biologically equivalent amino acids, as discussed in the following pages.

TABLE 1

Amino Acids			Codons
Alanine	Ala	A	GCA GCC GCG GCU
Cysteine	Cys	C	UGC UGU
Aspartic acid	Asp	D	GAC GAU
Glutamic acid	Glu	E	GAA GAG
Phenylalanine	Phe	F	UUC UUU
Glycine	Gly	G	GGA GGC GGG GGU
Histidine	His	H	CAC CAU
Isoleucine	Ile	I	AUA AUC AUU
Lysine	Lys	K	AAA AAG
Leucine	Leu	L	UUA UUG CUA CUC CUG CUU
Methionine	Met	M	AUG
Asparagine	Asn	N	AAC AAU
Proline	Pro	P	CCA CCC CCG CCU
Glutamine	Gln	Q	CAA CAG
Arginine	Arg	R	AGA AGG CGA CGC CGG CGU
Serine	Ser	S	AGC AGU UCA UCC UCG UCU
Threonine	Thr	T	ACA ACC ACG ACU
Valine	Val	V	GUA GUC GUG GUU
Tryptophan	Trp	W	UGG
Tyrosine	Tyr	Y	UAC UAU

Allowing for the degeneracy of the genetic code, sequences that have at least about 50%, usually at least about 60%, more usually about 70%, most usually about 80%, preferably at least about 90% and most preferably about 95% of nucleotides that are identical to the nucleotides of SEQ ID NO:1 will be sequences that are "as set forth in SEQ ID NO:1." Sequences that are essentially the same as those set forth in SEQ ID NO:1 may also be

functionally defined as sequences that are capable of hybridizing to a nucleic acid segment containing the complement of SEQ ID NO:1 under standard conditions.

The DNA segments of the present invention include those encoding biologically functional equivalent IEMT proteins and peptides, as described above. Such sequences may arise as a consequence of codon redundancy and amino acid functional equivalency that are known to occur naturally within nucleic acid sequences and the proteins thus encoded. Alternatively, functionally equivalent proteins or peptides may be created *via* the application of recombinant DNA technology, in which changes in the protein structure may be engineered, based on considerations of the properties of the amino acids being exchanged. Changes designed by man may be introduced through the application of site-directed mutagenesis techniques or may be introduced randomly and screened later for the desired function, as described below.

B. Oligonucleotide Probes and Primers

Naturally, the present invention also encompasses DNA segments that are complementary, or essentially complementary, to the sequence set forth in SEQ ID NO:1. Nucleic acid sequences that are "complementary" are those that are capable of base-pairing according to the standard Watson-Crick complementary rules. As used herein, the term "complementary sequences" means nucleic acid sequences that are substantially complementary, as may be assessed by the same nucleotide comparison set forth above, or as defined as being capable of hybridizing to the nucleic acid segment of SEQ ID NO:1 under relatively stringent conditions such as those described herein. Such sequences may encode the entire IEMT protein or functional or non-functional fragments thereof.

Alternatively, the hybridizing segments may be shorter oligonucleotides. Sequences of 17 bases long should occur only once in the genome and, therefore, suffice to specify a unique target sequence. Although shorter oligomers are easier to make and increase *in vivo* accessibility, numerous other factors are involved in determining the specificity of hybridization. Both binding affinity and sequence specificity of an oligonucleotide to its

complementary target increases with increasing length. It is contemplated that exemplary oligonucleotides of 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100 or more base pairs will be used, although others are contemplated. Longer polynucleotides encoding 250, 500, 1000, 1212, 1500, 2000, 2500, 3000 or 3431 bases and longer are contemplated as well. Such oligonucleotides will find use, for example, as probes in Southern and Northern blots and as primers in amplification reactions.

Suitable hybridization conditions will be well known to those of skill in the art. In certain applications, for example, substitution of amino acids by site-directed mutagenesis, it is appreciated that lower stringency conditions are required. Under these conditions, hybridization may occur even though the sequences of probe and target strand are not perfectly complementary, but are mismatched at one or more positions. Conditions may be rendered less stringent by increasing salt concentration and decreasing temperature. For example, a medium stringency condition could be provided by about 0.1 to 0.25 M NaCl at temperatures of about 37°C to about 55°C, while a low stringency condition could be provided by about 0.15 M to about 0.9 M salt, at temperatures ranging from about 20°C to about 55°C. Thus, hybridization conditions can be readily manipulated, and thus will generally be a method of choice depending on the desired results.

One method of using probes and primers of the present invention is in the search for genes related to IEMT or, more particularly, homologs of IEMT from other species. Normally, the target DNA will be a genomic or cDNA library, although screening may involve analysis of RNA molecules. By varying the stringency of hybridization, and the region of the probe, different degrees of homology may be discovered.

Another way of exploiting probes and primers of the present invention is in site-directed, or site-specific mutagenesis. Site-specific mutagenesis is a technique useful in the preparation of individual peptides, or biologically functional equivalent proteins or peptides, through specific mutagenesis of the underlying DNA. The technique further provides a ready ability to prepare and test sequence variants, incorporating one or more of the foregoing

considerations, by introducing one or more nucleotide sequence changes into the DNA. Site-specific mutagenesis allows the production of mutants through the use of specific oligonucleotide sequences which encode the DNA sequence of the desired mutation, as well as a sufficient number of adjacent nucleotides, to provide a primer sequence of sufficient size and sequence complexity to form a stable duplex on both sides of the deletion junction being traversed. Typically, a primer of about 17 to 25 nucleotides in length is preferred, with about 5 to 10 residues on both sides of the junction of the sequence being altered.

The technique typically employs a bacteriophage vector that exists in both a single-stranded and double-stranded form. Typical vectors useful in site-directed mutagenesis include vectors such as the M13 phage. These phage vectors are commercially available and their use is generally well known to those skilled in the art. Double stranded plasmids are also routinely employed in site directed mutagenesis, which eliminates the step of transferring the gene of interest from a phage to a plasmid.

In general, site-directed mutagenesis is performed by first obtaining a single-stranded vector, or melting of two strands of a double stranded vector which includes within its sequence a DNA sequence encoding the desired protein. An oligonucleotide primer bearing the desired mutated sequence is synthetically prepared. This primer is then annealed with the single-stranded DNA preparation, taking into account the degree of mismatch when selecting hybridization conditions, and subjected to DNA polymerizing enzymes such as *E. coli* polymerase I Klenow fragment, in order to complete the synthesis of the mutation-bearing strand. Thus, a heteroduplex is formed wherein one strand encodes the original non-mutated sequence and the second strand bears the desired mutation. This heteroduplex vector is then used to transform appropriate cells, such as *E. coli* cells, and clones are selected that include recombinant vectors bearing the mutated sequence arrangement.

Structure-guided site-specific mutagenesis represents a powerful tool for the dissection and engineering of protein-ligand interactions (Wells, 1996, Braisted *et al.*, 1996). One particularly useful mutagenesis technique is alanine scanning mutagenesis in which a number

of residues are substituted individually with the amino acid alanine so that the effects of losing side-chain interactions can be determined, while minimizing the risk of large-scale perturbations in protein conformation (Cunningham *et al.*, 1989).

Comprehensive information on the functional significance and information content of a given residue of protein can best be obtained by saturation mutagenesis in which all 19 amino acid substitutions are examined (Warren *et al.*, 1996, Brown *et al.*, 1996; Zeng *et al.*, 1996; Burton and Barbas, 1994; Yelton *et al.*, 1995; Jackson *et al.*, 1995; Short *et al.*, 1995; Wong *et al.*, 1996; Hilton *et al.*, 1996).

It is possible to bypass cloning steps by combining PCR mutagenesis with coupled *in vitro* transcription/translation for the high throughput generation of protein mutants (Burk et al., 1997). Here, the PCR products are used directly as the template for the *in vitro* transcription/translation of the mutant protein. Because of the high efficiency with which all 19 amino acid substitutions can be generated and analyzed in this way, it is now possible to perform saturation mutagenesis on numerous residues of interest, a process that can be described as *in vitro* scanning saturation mutagenesis (Burks *et al.*, 1997).

The preparation of sequence variants of the selected gene using site-directed mutagenesis is provided as a means of producing potentially useful species and is not meant to be limiting, as there are other ways in which sequence variants of genes may be obtained. For example, recombinant vectors encoding the desired gene may be treated with mutagenic agents, such as hydroxylamine, to obtain sequence variants.

C. Antisense Constructs

Antisense treatments are one way of inhibiting lignin biosynthesis in a plant. Antisense technology may be used to "knock-out" the function of the IEMT gene or other lignin biosynthesis genes, thereby decreasing or eliminating the expression of lignin precursors such as ferulic acid, coniferyl aldehyde or coniferyl alcohol a transformed plant cell or whole plant.

Antisense methodology takes advantage of the fact that nucleic acids tend to pair with "complementary" sequences. By complementary, it is meant that polynucleotides are those which are capable of base-pairing according to the standard Watson-Crick complementarity rules. That is, the larger purines will base pair with the smaller pyrimidines to form combinations of guanine paired with cytosine (G:C) and adenine paired with either thymine (A:T) in the case of DNA, or adenine paired with uracil (A:U) in the case of RNA. Inclusion of less common bases such as inosine, 5-methylcytosine, 6-methyladenine, hypoxanthine and others in hybridizing sequences does not interfere with pairing.

Targeting double-stranded (ds) DNA with polynucleotides leads to triple-helix formation; targeting RNA will lead to double-helix formation. Antisense polynucleotides, when introduced into a target cell, specifically bind to their target polynucleotide and interfere with transcription, RNA processing, transport, translation and/or stability. Antisense RNA constructs, or DNA encoding such antisense RNA's, may be employed to inhibit gene transcription or translation or both.

Antisense constructs may be designed to bind to the promoter and other control regions, exons, introns or even exon-intron boundaries of a gene. It is contemplated that the most effective antisense constructs will include regions complementary to intron/exon splice junctions. Thus, it is proposed that a preferred embodiment includes an antisense construct with complementarity to regions within 50-200 bases of an intron-exon splice junction. It has been observed that some exon sequences can be included in the construct without seriously affecting the target selectivity thereof. The amount of exonic material included will vary depending on the particular exon and intron sequences used. One can readily test whether too much exon DNA is included simply by testing the constructs *in vitro* to determine whether normal cellular function is affected or whether the expression of related genes having complementary sequences is affected.

As stated above, "complementary" or "antisense" means polynucleotide sequences that are substantially complementary over their entire length and have very few base mismatches.

For example, sequences of fifteen bases in length may be termed complementary when they have complementary nucleotides at thirteen or fourteen positions. Naturally, sequences which are completely complementary will be sequences which are entirely complementary throughout their entire length and have no base mismatches. Other sequences with lower degrees of homology also are contemplated. For example, an antisense construct which has limited regions of high homology, but also contains a non-homologous region (*e.g.*, ribozyme; see below) could be designed. These molecules, though having less than 50% homology, would bind to target sequences under appropriate conditions.

It may be advantageous to combine portions of genomic DNA with cDNA or synthetic sequences to generate specific constructs. For example, where an intron is desired in the ultimate construct, a genomic clone will need to be used. The cDNA or a synthesized polynucleotide may provide more convenient restriction sites for the remaining portion of the construct and, therefore, would be used for the rest of the sequence.

IV. Generating Antibodies Reactive With IEMT Proteins

In another aspect, the present invention contemplates an antibody that is immunoreactive with a IEMT molecule of the present invention, or any portion thereof. An antibody can be a polyclonal or a monoclonal antibody. In a preferred embodiment, an antibody is a monoclonal antibody. Means for preparing and characterizing antibodies are well known in the art (see, *e.g.*, Howell and Lane, 1988).

In a particular embodiment, one may exploit antibodies to either IEMT or to COMT as a way of distinguishing various chimeric constructs. In particular, antibodies that recognize epitopes on these antigens that are lost or disturbed in chimeras may be used to identify IEMT or COMT polypeptides while not reacting with the chimeras. Conversely, the chimeras may result in new epitopes not present in either IEMT or COMT. Thus, these antibodies will prove useful in examining expression levels of cells and in selection of cells where the cells express an IEMT/COMT chimera and at least one of IEMT and COMT.

Briefly, a polyclonal antibody is prepared by immunizing an animal with an immunogen comprising a polypeptide of the present invention and collecting antisera from that immunized animal. A wide range of animal species can be used for the production of antisera. Typically an animal used for production of anti-antisera is a non-human animal including rabbits, mice, rats, hamsters, pigs or horses. Because of the relatively large blood volume of rabbits, a rabbit is a preferred choice for production of polyclonal antibodies.

Antibodies, both polyclonal and monoclonal, specific for isoforms of antigen may be prepared using conventional immunization techniques, as will be generally known to those of skill in the art. A composition containing antigenic epitopes of the compounds of the present invention can be used to immunize one or more experimental animals, such as a rabbit or mouse, which will then proceed to produce specific antibodies against the compounds of the present invention. Polyclonal antisera may be obtained, after allowing time for antibody generation, simply by bleeding the animal and preparing serum samples from the whole blood.

It is proposed that the monoclonal antibodies of the present invention will find useful application in standard immunochemical procedures, such as ELISA and Western blot methods and in immunohistochemical procedures such as tissue staining, as well as in other procedures which may utilize antibodies specific to IEMT-related antigen epitopes. Additionally, it is proposed that monoclonal antibodies specific to a particular IEMT protein of different species may be utilized in other useful applications.

In general, both polyclonal and monoclonal antibodies against IEMT may be used in a variety of embodiments. For example, they may be employed in antibody cloning protocols to obtain cDNAs or genes encoding other IEMT proteins. They may also be used in inhibition studies to analyze the effects of IEMT related peptides in cells or plants. Anti-IEMT antibodies will also be useful in immunolocalization studies to analyze the distribution of IEMT during various cellular events or stages of development. A particularly useful application of such antibodies is in purifying native or recombinant IEMT polypeptides, for example, using an

antibody affinity column. The operation of all such immunological techniques will be known to those of skill in the art in light of the present disclosure.

Means for preparing and characterizing antibodies are well known in the art (see, *e.g.*, Harlow and Lane, 1988; incorporated herein by reference). More specific examples of monoclonal antibody preparation are given in the examples below.

As is well known in the art, a given composition may vary in its immunogenicity. It is often necessary therefore to boost the host immune system, as may be achieved by coupling a peptide or polypeptide immunogen to a carrier. Exemplary and preferred carriers are keyhole limpet hemocyanin (KLH) and bovine serum albumin (BSA). Other albumins such as ovalbumin, mouse serum albumin or rabbit serum albumin can also be used as carriers. Means for conjugating a polypeptide to a carrier protein are well known in the art and include glutaraldehyde, *m*-maleimidobencoyl-N-hydroxysuccinimide ester, carbodiimide and bis-biazotized benzidine.

As also is well known in the art, the immunogenicity of a particular immunogen composition can be enhanced by the use of non-specific stimulators of the immune response, known as adjuvants. Exemplary and preferred adjuvants include complete Freund's adjuvant (a non-specific stimulator of the immune response containing killed *Mycobacterium tuberculosis)*, incomplete Freund's adjuvants and aluminum hydroxide adjuvant.

The amount of immunogen composition used in the production of polyclonal antibodies varies upon the nature of the immunogen as well as the animal used for immunization. A variety of routes can be used to administer the immunogen (subcutaneous, intramuscular, intradermal, intravenous and intraperitoneal). The production of polyclonal antibodies may be monitored by sampling blood of the immunized animal at various points following immunization. A second, booster, injection may also be given. The process of boosting and titering is repeated until a suitable titer is achieved. When a desired level of immunogenicity is

obtained, the immunized animal can be bled and the serum isolated and stored, and/or the animal can be used to generate mAbs.

MAbs may be readily prepared through use of well-known techniques, such as those exemplified in U.S. Patent 4,196,265, incorporated herein by reference. Typically, this technique involves immunizing a suitable animal with a selected immunogen composition, *e.g.*, a purified or partially purified IEMT protein, polypeptide or peptide or cell expressing high levels of IEMT. The immunizing composition is administered in a manner effective to stimulate antibody producing cells. Rodents such as mice and rats are preferred animals, however, the use of rabbit, sheep frog cells is also possible. The use of rats may provide certain advantages (Goding, 1986), but mice are preferred, with the BALB/c mouse being most preferred as this is most routinely used and generally gives a higher percentage of stable fusions.

Following immunization, somatic cells with the potential for producing antibodies, specifically B-lymphocytes (B-cells), are selected for use in the mAb generating protocol. These cells may be obtained from biopsied spleens, tonsils or lymph nodes, or from a peripheral blood sample. Spleen cells and peripheral blood cells are preferred, the former because they are a rich source of antibody-producing cells that are in the dividing plasmablast stage, and the latter because peripheral blood is easily accessible. Often, a panel of animals will have been immunized and the spleen of animal with the highest antibody titer will be removed and the spleen lymphocytes obtained by homogenizing the spleen with a syringe. Typically, a spleen from an immunized mouse contains approximately 5×10^7 to 2×10^8 lymphocytes.

The antibody-producing B lymphocytes from the immunized animal are then fused with cells of an immortal myeloma cell, generally one of the same species as the animal that was immunized. Myeloma cell lines suited for use in hybridoma-producing fusion procedures preferably are non-antibody-producing, have high fusion efficiency, and enzyme deficiencies that render then incapable of growing in certain selective media which support the growth of only the desired fused cells (hybridomas).

Any one of a number of myeloma cells may be used, as are known to those of skill in the art (Goding, 1986; Campbell, 1984). For example, where the immunized animal is a mouse, one may use P3-X63/Ag8, P3-X63-Ag8.653, NS1/1.Ag 4 1, Sp210-Ag14, FO, NSO/U, MPC-11, MPC11-X45-GTG 1.7 and S194/5XX0 Bul; for rats, one may use R210.RCY3, Y3-Ag 1.2.3, IR983F and 4B210; and U-266, GM1500-GRG2, LICR-LON-HMy2 and UC729-6 are all useful in connection with cell fusions.

Methods for generating hybrids of antibody-producing spleen or lymph node cells and myeloma cells usually comprise mixing somatic cells with myeloma cells in a 2:1 ratio, though the ratio may vary from about 20:1 to about 1:1, respectively, in the presence of an agent or agents (chemical or electrical) that promote the fusion of cell membranes. Fusion methods using Sendai virus have been described (Kohler and Milstein, 1975; 1976), and those using polyethylene glycol (PEG), such as 37% (v/v) PEG, by Gefter *et al.*, (1977). The use of electrically induced fusion methods is also appropriate (Goding, 1986).

Fusion procedures usually produce viable hybrids at low frequencies, around 1×10^{-6} to 1×10^{-8}. However, this does not pose a problem, as the viable, fused hybrids are differentiated from the parental, unfused cells (particularly the unfused myeloma cells that would normally continue to divide indefinitely) by culturing in a selective medium. The selective medium is generally one that contains an agent that blocks the *de novo* synthesis of nucleotides in the tissue culture media. Exemplary and preferred agents are aminopterin, methotrexate, and azaserine. Aminopterin and methotrexate block *de novo* synthesis of both purines and pyrimidines, whereas azaserine blocks only purine synthesis. Where aminopterin or methotrexate is used, the media is supplemented with hypoxanthine and thymidine as a source of nucleotides (HAT medium). Where azaserine is used, the media is supplemented with hypoxanthine.

The preferred selection medium is HAT. Only cells capable of operating nucleotide salvage pathways are able to survive in HAT medium. The myeloma cells are defective in key

enzymes of the salvage pathway, *e.g.*, hypoxanthine phosphoribosyl transferase (HPRT), and they cannot survive. The B-cells can operate this pathway, but they have a limited life span in culture and generally die within about two wk. Therefore, the only cells that can survive in the selective media are those hybrids formed from myeloma and B-cells.

This culturing provides a population of hybridomas from which specific hybridomas are selected. Typically, selection of hybridomas is performed by culturing the cells by single-clone dilution in microtiter plates, followed by testing the individual clonal supernatants (after about two to three wk) for the desired reactivity. The assay should be sensitive, simple and rapid, such as radioimmunoassays, enzyme immunoassays, cytotoxicity assays, plaque assays, dot immunobinding assays, and the like.

The selected hybridomas would then be serially diluted and cloned into individual antibody-producing cell lines, which clones can then be propagated indefinitely to provide mAbs. The cell lines may be exploited for mAb production in two basic ways. A sample of the hybridoma can be injected (often into the peritoneal cavity) into a histocompatible animal of the type that was used to provide the somatic and myeloma cells for the original fusion. The injected animal develops tumors secreting the specific monoclonal antibody produced by the fused cell hybrid. The body fluids of the animal, such as serum or ascites fluid, can then be tapped to provide mAbs in high concentration. The individual cell lines could also be cultured *in vitro*, where the mAbs are naturally secreted into the culture medium from which they can be readily obtained in high concentrations. mAbs produced by either means may be further purified, if desired, using filtration, centrifugation and various chromatographic methods such as HPLC or affinity chromatography.

V. Genetic Analysis of IEMT Transgenic Plants

One embodiment of the instant invention comprises a method for detecting variation in the expression of IEMT genes. As used herein, the term "IEMT gene" is meant to represent a gene of phenylpropanoid biosynthesis which includes IEMT, and other methyl transferases

capable of catalyzing the 4 hydroxy methylation of compounds of the lignin biosynthetic pathway. This method may comprise determining that level of IEMT protein or determining specific alterations in the expressed product. Obviously, this sort of assay has importance in the screening of transformants for potential changes in flavor, fragrance and lignification. Such assays may in some cases be faster, more accurate or less expensive than conventional assays.

The biological sample may potentially be any type of plant tissue. Nucleic acid is isolated from cells contained in the biological sample, according to standard methodologies (Sambrook *et al.*, 1989). The nucleic acid may be genomic DNA or fractionated or whole cell RNA. Where RNA is used, it may be desired to convert the RNA to a complementary DNA. In one embodiment, the RNA is whole cell RNA; in another, it is poly-A RNA. Normally, the nucleic acid is amplified.

Depending on the format, the specific nucleic acid of interest is identified in the sample directly using amplification or with a second, known nucleic acid following amplification. Next, the identified product is detected. In certain applications, the detection may be performed by visual means (*e.g.*, ethidium bromide staining of a gel). Alternatively, the detection may involve indirect identification of the product *via* chemiluminescence, radioactive scintigraphy of radiolabel or fluorescent label or even *via* a system using electrical or thermal impulse signals (Affymax Technology; Bellus, 1994).

Following detection, one may compare the results seen in a given plant with a statistically significant reference group of non-transformed control plants. Typically, the non-transformed control plants will be of a genetic background similar to the transformed plants. In this way, it is possible to detect differences in the amount or kind of IEMT protein detected in various transformed plants.

A variety of different assays are contemplated in the screening of plants for particular IEMT transgenes and associated exogenous elements. These techniques may in cases be used to detect for both the presence of the particular genes as well as rearrangements that may have

occurred in the gene construct. The techniques include but are not limited to, fluorescent *in situ* hybridization (FISH), direct DNA sequencing, PFGE analysis, Southern or Northern blotting, single-stranded conformation analysis (SSCA), RNAse protection assay, allele-specific oligonucleotide (ASO), dot blot analysis, denaturing gradient gel electrophoresis, RFLP and PCR-SSCP.

A. Primers and Probes

The term primer, as defined herein, is meant to encompass any nucleic acid that is capable of priming the synthesis of a nascent nucleic acid in a template-dependent process. Typically, primers are oligonucleotides from ten to twenty base pairs in length, but longer sequences can be employed. Primers may be provided in double-stranded or single-stranded form, although the single-stranded form is preferred. Probes are defined differently, although they may act as primers. Probes, while perhaps capable of priming, are designed to binding to the target DNA or RNA and need not be used in an amplification process.

In preferred embodiments, the probes or primers are labeled with radioactive species (^{32}P, ^{14}C, ^{35}S, ^{3}H, or other label), with a fluorophore (rhodamine, fluorescein), an antigen (biotin, streptavidin, digoxigenin), or a chemillumiscent (luciferase).

B. Template Dependent Amplification Methods

A number of template dependent processes are available to amplify the marker sequences present in a given template sample. One of the best known amplification methods is the polymerase chain reaction (referred to as PCR™) which is described in detail in U.S. Patent Nos. 4,683,195, 4,683,202 and 4,800,159, and in Innis *et al.*, 1990, each of which is incorporated herein by reference in its entirety.

Briefly, in PCR, two primer sequences are prepared that are complementary to regions on opposite complementary strands of the marker sequence. An excess of deoxynucleoside triphosphates are added to a reaction mixture along with a DNA polymerase, *e.g.*, *Taq* polymerase. If the marker sequence is present in a sample, the primers will bind to the marker

and the polymerase will cause the primers to be extended along the marker sequence by adding on nucleotides. By raising and lowering the temperature of the reaction mixture, the extended primers will dissociate from the marker to form reaction products, excess primers will bind to the marker and to the reaction products and the process is repeated.

A reverse transcriptase PCR amplification procedure may be performed in order to quantify the amount of mRNA amplified. Methods of reverse transcribing RNA into cDNA are well known and described in Sambrook *et al.*, 1989. Alternative methods for reverse transcription utilize thermostable, RNA-dependent DNA polymerases. These methods are described in WO 90/07641 filed December 21, 1990. Polymerase chain reaction methodologies are well known in the art.

Another method for amplification is the ligase chain reaction ("LCR"), disclosed in EPO No. 320 308, incorporated herein by reference in its entirety. In LCR, two complementary probe pairs are prepared, and in the presence of the target sequence, each pair will bind to opposite complementary strands of the target such that they abut. In the presence of a ligase, the two probe pairs will link to form a single unit. By temperature cycling, as in PCR, bound ligated units dissociate from the target and then serve as "target sequences" for ligation of excess probe pairs. U.S. Patent 4,883,750 describes a method similar to LCR for binding probe pairs to a target sequence.

Qbeta Replicase, described in PCT Application No. PCT/US87/00880, may also be used as still another amplification method in the present invention. In this method, a replicative sequence of RNA that has a region complementary to that of a target is added to a sample in the presence of an RNA polymerase. The polymerase will copy the replicative sequence that can then be detected.

An isothermal amplification method, in which restriction endonucleases and ligases are used to achieve the amplification of target molecules that contain nucleotide 5'-[alpha-thio]-

triphosphates in one strand of a restriction site may also be useful in the amplification of nucleic acids in the present invention, Walker *et al.*, (1992).

Strand Displacement Amplification (SDA) is another method of carrying out isothermal amplification of nucleic acids which involves multiple rounds of strand displacement and synthesis, *i.e.*, nick translation. A similar method, called Repair Chain Reaction (RCR), involves annealing several probes throughout a region targeted for amplification, followed by a repair reaction in which only two of the four bases are present. The other two bases can be added as biotinylated derivatives for easy detection. A similar approach is used in SDA. Target specific sequences can also be detected using a cyclic probe reaction (CPR). In CPR, a probe having 3' and 5' sequences of non-specific DNA and a middle sequence of specific RNA is hybridized to DNA that is present in a sample. Upon hybridization, the reaction is treated with RNase H, and the products of the probe identified as distinctive products that are released after digestion. The original template is annealed to another cycling probe and the reaction is repeated.

Still another amplification methods described in GB Application No. 2 202 328, and in PCT Application No. PCT/US89/01025, each of which is incorporated herein by reference in its entirety, may be used in accordance with the present invention. In the former application, "modified" primers are used in a PCR-like, template- and enzyme-dependent synthesis. The primers may be modified by labeling with a capture moiety (*e.g.*, biotin) and/or a detector moiety (*e.g.*, enzyme). In the latter application, an excess of labeled probes are added to a sample. In the presence of the target sequence, the probe binds and is cleaved catalytically. After cleavage, the target sequence is released intact to be bound by excess probe. Cleavage of the labeled probe signals the presence of the target sequence.

Other nucleic acid amplification procedures include transcription-based amplification systems (TAS), including nucleic acid sequence based amplification (NASBA) and 3SR (Kwoh *et al.*, 1989; Gingeras *et al.*, PCT Application WO 88/10315, incorporated herein by reference in their entirety). In NASBA, the nucleic acids can be prepared for amplification by standard

phenol/chloroform extraction, heat denaturation of a clinical sample, treatment with lysis buffer and minispin columns for isolation of DNA and RNA or guanidinium chloride extraction of RNA. These amplification techniques involve annealing a primer which has target specific sequences. Following polymerization, DNA/RNA hybrids are digested with RNase H while double stranded DNA molecules are heat denatured again. In either case the single stranded DNA is made fully double stranded by addition of second target specific primer, followed by polymerization. The double-stranded DNA molecules are then multiply transcribed by an RNA polymerase such as T7 or SP6. In an isothermal cyclic reaction, the RNA's are reverse transcribed into single stranded DNA, which is then converted to double stranded DNA, and then transcribed once again with an RNA polymerase such as T7 or SP6. The resulting products, whether truncated or complete, indicate target specific sequences.

Davey *et al.*, EPO No. 329 822 (incorporated herein by reference in its entirety) disclose a nucleic acid amplification process involving cyclically synthesizing single-stranded RNA ("ssRNA"), ssDNA, and double-stranded DNA (dsDNA), which may be used in accordance with the present invention. The ssRNA is a template for a first primer oligonucleotide, which is elongated by reverse transcriptase (RNA-dependent DNA polymerase). The RNA is then removed from the resulting DNA:RNA duplex by the action of ribonuclease H (RNase H, an RNase specific for RNA in duplex with either DNA or RNA). The resultant ssDNA is a template for a second primer, which also includes the sequences of an RNA polymerase promoter (exemplified by T7 RNA polymerase) 5' to its homology to the template. This primer is then extended by DNA polymerase (exemplified by the large "Klenow" fragment of *E. coli* DNA polymerase I), resulting in a double-stranded DNA ("dsDNA") molecule, having a sequence identical to that of the original RNA between the primers and having additionally, at one end, a promoter sequence. This promoter sequence can be used by the appropriate RNA polymerase to make many RNA copies of the DNA. These copies can then re-enter the cycle leading to very swift amplification. With proper choice of enzymes, this amplification can be done isothermally without addition of enzymes at each cycle. Because of the cyclical nature of this process, the starting sequence can be chosen to be in the form of either DNA or RNA.

Miller *et al.*, PCT Application WO 89/06700 (incorporated herein by reference in its entirety) disclose a nucleic acid sequence amplification scheme based on the hybridization of a promoter/primer sequence to a target single-stranded DNA ("ssDNA") followed by transcription of many RNA copies of the sequence. This scheme is not cyclic, *i.e.*, new templates are not produced from the resultant RNA transcripts. Other amplification methods include "RACE" and "one-sided PCR" (Frohman, M.A., In: *PCR PROTOCOLS: A GUIDE TO METHODS AND APPLICATIONS*, Academic Press, N.Y., 1990; Ohara *et al.*, 1989; each herein incorporated by reference in their entirety).

Methods based on ligation of two (or more) oligonucleotides in the presence of nucleic acid having the sequence of the resulting "di-oligonucleotide", thereby amplifying the di-oligonucleotide, may also be used in the amplification step of the present invention. Wu *et al.*, (1989), incorporated herein by reference in its entirety.

C. Southern/Northern Blotting

Blotting techniques are well known to those of skill in the art. Southern blotting involves the use of DNA as a target, whereas Northern blotting involves the use of RNA as a target. Each provide different types of information, although cDNA blotting is analogous, in many aspects, to blotting or RNA species.

Briefly, a probe is used to target a DNA or RNA species that has been immobilized on a suitable matrix, often a filter of nitrocellulose. The different species should be spatially separated to facilitate analysis. This often is accomplished by gel electrophoresis of nucleic acid species followed by "blotting" on to the filter.

Subsequently, the blotted target is incubated with a probe (usually labeled) under conditions that promote denaturation and rehybridization. Because the probe is designed to base pair with the target, the probe will binding a portion of the target sequence under renaturing conditions. Unbound probe is then removed, and detection is accomplished as described above.

D. Separation Methods

It normally is desirable, at one stage or another, to separate the amplification product from the template and the excess primer for the purpose of determining whether specific amplification has occurred. In one embodiment, amplification products are separated by agarose, agarose-acrylamide or polyacrylamide gel electrophoresis using standard methods. See Sambrook *et al.*, 1989.

Alternatively, chromatographic techniques may be employed to effect separation. There are many kinds of chromatography which may be used in the present invention: adsorption, partition, ion-exchange and molecular sieve, and many specialized techniques for using them including column, paper, thin-layer and gas chromatography (Freifelder, 1982).

E. Detection Methods

Products may be visualized in order to confirm amplification of the marker sequences. One typical visualization method involves staining of a gel with ethidium bromide and visualization under UV light. Alternatively, if the amplification products are integrally labeled with radio- or fluorometrically-labeled nucleotides, the amplification products can then be exposed to x-ray film or visualized under the appropriate stimulating spectra, following separation.

In one embodiment, visualization is achieved indirectly. Following separation of amplification products, a labeled nucleic acid probe is brought into contact with the amplified marker sequence. The probe preferably is conjugated to a chromophore but may be radiolabeled. In another embodiment, the probe is conjugated to a binding partner, such as an antibody or biotin, and the other member of the binding pair carries a detectable moiety.

In one embodiment, detection is by a labeled probe. The techniques involved are well known to those of skill in the art and can be found in many standard books on molecular

protocols. See Sambrook *et al.*, 1989. For example, chromophore or radiolabel probes or primers identify the target during or following amplification.

One example of the foregoing is described in U.S. Patent No. 5,279,721, incorporated by reference herein, which discloses an apparatus and method for the automated electrophoresis and transfer of nucleic acids. The apparatus permits electrophoresis and blotting without external manipulation of the gel and is ideally suited to carrying out methods according to the present invention.

In addition, the amplification products described above may be subjected to sequence analysis to identify specific kinds of variations using standard sequence analysis techniques. Within certain methods, exhaustive analysis of genes is carried out by sequence analysis using primer sets designed for optimal sequencing (Pignon *et al*, 1994). The present invention provides methods by which any or all of these types of analyses may be used. Using the sequences disclosed herein, oligonucleotide primers may be designed to permit the amplification of sequences throughout the IEMT1 or other IEMT genes that may then be analyzed by direct sequencing.

F. Kit Components

All the essential materials and reagents required for detecting and sequencing IEMT1 or other IEMT genes, and variants thereof may be assembled together in a kit. This generally will comprise preselected primers and probes. Also included may be enzymes suitable for amplifying nucleic acids including various polymerases (RT, Taq, Sequenase™ *etc.*), deoxynucleotides and buffers to provide the necessary reaction mixture for amplification. Such kits also generally will comprise, in suitable means, distinct containers for each individual reagent and enzyme as well as for each primer or probe.

G. Design and Theoretical Considerations for Relative Quantitative RT-PCR

Reverse transcription (RT) of RNA to cDNA followed by relative quantitative PCR (RT-PCR) can be used to determine the relative concentrations of specific mRNA species

isolated from plants. By determining that the concentration of a specific mRNA species varies, it is shown that the gene encoding the specific mRNA species is differentially expressed.

In PCR, the number of molecules of the amplified target DNA increase by a factor approaching two with every cycle of the reaction until some reagent becomes limiting. Thereafter, the rate of amplification becomes increasingly diminished until there is no increase in the amplified target between cycles. If a graph is plotted in which the cycle number is on the X axis and the log of the concentration of the amplified target DNA is on the Y axis, a curved line of characteristic shape is formed by connecting the plotted points. Beginning with the first cycle, the slope of the line is positive and constant. This is said to be the linear portion of the curve. After a reagent becomes limiting, the slope of the line begins to decrease and eventually becomes zero. At this point the concentration of the amplified target DNA becomes asymptotic to some fixed value. This is said to be the plateau portion of the curve.

The concentration of the target DNA in the linear portion of the PCR amplification is directly proportional to the starting concentration of the target before the reaction began. By determining the concentration of the amplified products of the target DNA in PCR reactions that have completed the same number of cycles and are in their linear ranges, it is possible to determine the relative concentrations of the specific target sequence in the original DNA mixture. If the DNA mixtures are cDNAs synthesized from RNAs isolated from different tissues or cells, the relative abundances of the specific mRNA from which the target sequence was derived can be determined for the respective tissues or cells. This direct proportionality between the concentration of the PCR products and the relative mRNA abundances is only true in the linear range of the PCR reaction.

The final concentration of the target DNA in the plateau portion of the curve is determined by the availability of reagents in the reaction mix and is independent of the original concentration of target DNA. Therefore, the first condition that must be met before the relative abundances of a mRNA species can be determined by RT-PCR for a collection of RNA

populations is that the concentrations of the amplified PCR products must be sampled when the PCR reactions are in the linear portion of their curves.

The second condition that must be met for an RT-PCR study to successfully determine the relative abundances of a particular mRNA species is that relative concentrations of the amplifiable cDNAs must be normalized to some independent standard. The goal of an RT-PCR study is to determine the abundance of a particular mRNA species relative to the average abundance of all mRNA species in the sample.

Most protocols for competitive PCR utilize internal PCR standards that are approximately as abundant as the target. These strategies are effective if the products of the PCR amplifications are sampled during their linear phases. If the products are sampled when the reactions are approaching the plateau phase, then the less abundant product becomes relatively over represented. Comparisons of relative abundances made for many different RNA samples, such as is the case when examining RNA samples for differential expression, become distorted in such a way as to make differences in relative abundances of RNAs appear less than they actually are. This is not a significant problem if the internal standard is much more abundant than the target. If the internal standard is more abundant than the target, then direct linear comparisons can be made between RNA samples.

The above discussion describes theoretical considerations for an RT-PCR assay for plant tissue. The problems inherent in plant tissue samples are that they are of variable quantity (making normalization problematic), and that they are of variable quality (necessitating the co-amplification of a reliable internal control, preferably of larger size than the target). Both of these problems are overcome if the RT-PCR is performed as a relative quantitative RT-PCR with an internal standard in which the internal standard is an amplifiable cDNA fragment that is larger than the target cDNA fragment and in which the abundance of the mRNA encoding the internal standard is roughly 5-100 fold higher than the mRNA encoding the target. This assay measures relative abundance, not absolute abundance of the respective mRNA species.

Other studies may be performed using a more conventional relative quantitative RT-PCR assay with an external standard protocol. These assays sample the PCR products in the linear portion of their amplification curves. The number of PCR cycles that are optimal for sampling must be empirically determined for each target cDNA fragment. In addition, the reverse transcriptase products of each RNA population isolated from the various tissue samples must be carefully normalized for equal concentrations of amplifiable cDNAs. This consideration is very important since the assay measures absolute mRNA abundance. Absolute mRNA abundance can be used as a measure of differential gene expression only in normalized samples. While empirical determination of the linear range of the amplification curve and normalization of cDNA preparations are tedious and time consuming processes, the resulting RT-PCR assays can be superior to those derived from the relative quantitative RT-PCR assay with an internal standard.

One reason for this advantage is that without the internal standard/competitor, all of the reagents can be converted into a single PCR product in the linear range of the amplification curve, thus increasing the sensitivity of the assay. Another reason is that with only one PCR product, display of the product on an electrophoretic gel or another display method becomes less complex, has less background and is easier to interpret.

H. Chip Technologies

Specifically contemplated by the present inventors are chip-based DNA technologies such as those described by Hacia *et al.* (1996) and Shoemaker *et al.* (1996). Briefly, these techniques involve quantitative methods for analyzing large numbers of genes rapidly and accurately. By tagging genes with oligonucleotides or using fixed probe arrays, one can employ chip technology to segregate target molecules as high density arrays and screen these molecules on the basis of hybridization. See also Pease *et al.* (1994); Fodor *et al.* (1991).

VI. Immunoassays for IEMT Gene Expression

Antibodies of the present invention can be used in characterizing the expression of IEMT genes, through techniques such as ELISAs and Western blotting. This may provide a more efficient, accurate or cost effective method for screening plants as part of a standard breeding program or that have been transformed with IEMT genes for relative rates of phenylpropanoid biosynthesis, aiding in the selection of plant products that have a reduced clove-like flavor.

The use of antibodies of the present invention, in an ELISA assay is contemplated. For example, anti-IEMT proteins are immobilized onto a selected surface, preferably a surface exhibiting a protein affinity such as the wells of a polystyrene microtiter plate. After washing to remove incompletely adsorbed material, it is desirable to bind or coat the assay plate wells with a non-specific protein that is known to be antigenically neutral with regard to the test antisera such as bovine serum albumin (BSA), casein or solutions of powdered milk. This allows for blocking of non-specific adsorption sites on the immobilizing surface and thus reduces the background caused by non-specific binding of antigen onto the surface.

After binding of antibody to the well, coating with a non-reactive material to reduce background, and washing to remove unbound material, the immobilizing surface is contacted with the sample to be tested in a manner conducive to immune complex (antigen/antibody) formation.

Following formation of specific immunocomplexes between the test sample and the bound antibody, and subsequent washing, the occurrence and even amount of immunocomplex formation may be determined by subjecting same to a second antibody having specificity for IEMT proteins that differs from the first antibody. Appropriate conditions preferably include diluting the sample with diluents such as BSA, bovine gamma globulin (BGG) and phosphate buffered saline (PBS)/Tween®. These added agents also tend to assist in the reduction of nonspecific background. The layered antisera is then allowed to incubate for from about 2 to

about 4 hr, at temperatures preferably on the order of about 25° to about 27°C. Following incubation, the antisera-contacted surface is washed so as to remove non-immunocomplexed material. A preferred washing procedure includes washing with a solution such as PBS/Tween®, or borate buffer.

To provide a detecting means, the second antibody will preferably have an associated enzyme that will generate a color development upon incubating with an appropriate chromogenic substrate. Thus, for example, one will desire to contact and incubate the second antibody-bound surface with a urease or peroxidase-conjugated anti-human IgG for a period of time and under conditions which favor the development of immunocomplex formation (*e.g.*, incubation for 2 hr at room temperature in a PBS-containing solution such as PBS/Tween®).

After incubation with the second enzyme-tagged antibody, and subsequent to washing to remove unbound material, the amount of label is quantified by incubation with a chromogenic substrate such as urea and bromocresol purple or 2,2'-azino-di-(3-ethyl-benzthiazoline)-6-sulfonic acid (ABTS) and H_2O_2, in the case of peroxidase as the enzyme label. Quantitation is then achieved by measuring the degree of color generation, *e.g.*, using a visible spectrum spectrophotometer.

The preceding format may be altered by first binding the sample to the assay plate. Then, primary antibody is incubated with the assay plate, followed by detecting of bound primary antibody using a labeled second antibody with specificity for the primary antibody.

VII. Expression Vectors

The present invention contemplates an expression vector comprising a polynucleotide of the present invention. Thus, in one embodiment an expression vector is an isolated and purified DNA molecule comprising a promoter operatively linked to a coding region that encodes a polypeptide of the present invention, which coding region is operatively linked to a

transcription-terminating region, whereby the promoter drives the transcription of the coding region.

As used herein, the term "operatively linked" means that a promoter is connected to an coding region in such a way that the transcription of that coding region is controlled and regulated by that promoter. Means for operatively linking a promoter to a coding region are well known in the art.

In a preferred embodiment, the recombinant expression of DNAs encoding the IEMT proteins of the present invention is preferable in gram-negative bacterium such as an *E. coli* or *Pseudomonas* spp. host cell. Promoters which function in high-level expression of target polypeptides in *E. coli* and other Gram-negative host cells are well-known in the art. Alternatively, mutagenized or recombinant IEMT protein-encoding gene promoters may be engineered by the hand of man and used to promote expression of the novel gene segments disclosed herein.

Where an expression vector of the present invention is to be used to transform a plant, a promoter is selected that has the ability to drive expression in plants. Promoters that function in plants are also well known in the art. Useful in expressing the polypeptide in plants are promoters that are inducible, viral, synthetic, constitutive as described (Poszkowski *et al.*, 1989; Odell *et al.*, 1985), and temporally regulated, spatially regulated, and spatio-temporally regulated (Chau *et al.*, 1989).

A promoter is also selected for its ability to direct the transformed plant cell's or transgenic plant's transcriptional activity to the coding region. Structural genes can be driven by a variety of promoters in plant tissues. Promoters can be near-constitutive, such as the CaMV 35S promoter, actin promoter, histone promoters, and ubiquitin promoters or tissue-specific or developmentally specific promoters affecting dicots or monocots. Such promoters are discussed in U.S. Pat. Appl. No. 08/113,561, filed August 25, 1993, which is specifically incorporated herein by reference.

Where the promoter is a near-constitutive promoter such as CaMV 35S, increases in polypeptide expression are found in a variety of transformed plant tissues (*e.g.*, callus, leaf, seed and root). Alternatively, the effects of transformation can be directed to specific plant tissues by using plant integrating vectors containing a tissue-specific promoter.

An exemplary tissue-specific promoter is the lectin promoter, which is specific for seed tissue. The Lectin protein in soybean seeds is encoded by a single gene (Le1) that is only expressed during seed maturation and accounts for about 2 to about 5% of total seed mRNA. The lectin gene and seed-specific promoter have been fully characterized and used to direct seed specific expression in transgenic tobacco plants (Vodkin *et al.*, 1983; Lindstrom *et al.*, 1990.). Similar are the maize zein and globulin-1 promoters.

An expression vector containing a coding region that encodes a polypeptide of interest is engineered to be under control of the lectin promoter and that vector is introduced into plants using, for example, a protoplast transformation method (Dhir *et al.*, 1991) or microprojectile bombardment (U.S. Pat. Appl. No. 08/113,561, filed August 25, 1993). The expression of the polypeptide is directed specifically to the seeds of the transgenic plant.

A transgenic plant of the present invention produced from a plant cell transformed with a tissue specific promoter can be crossed with a second transgenic plant developed from a plant cell transformed with a different tissue specific promoter to produce a hybrid transgenic plant that shows the effects of transformation in more than one specific tissue.

Exemplary tissue-specific promoters are corn sucrose synthetase 1 (Yang *et al.*, 1990), corn alcohol dehydrogenase 1 (Vogel *et al.*, 1989), corn light harvesting complex (Simpson, 1986), corn heat shock protein (Odell *et al.*, 1985), pea small subunit RuBP carboxylase (Poulsen *et al.*, 1986; Cashmore *et al.*, 1983), Ti plasmid mannopine synthase (Langridge *et al.*, 1989), Ti plasmid nopaline synthase (Langridge *et al.*, 1989), petunia chalcone isomerase (Van Tunen *et al.*, 1988), bean glycine rich protein 1, CaMV 35s transcript (Odell *et al.*, 1985) and

Potato patatin (Wenzler *et al.*, 1989). Preferred promoters are the cauliflower mosaic virus (CaMV 35S) promoter and the rice actin promoter.

The choice of which expression vector and ultimately to which promoter a polypeptide coding region is operatively linked depends directly on the functional properties desired, *e.g.*, the location and timing of protein expression, and the host cell to be transformed. These are well known limitations inherent in the art of constructing recombinant DNA molecules. However, a vector useful in practicing the present invention is capable of directing the expression of the polypeptide coding region to which it is operatively linked.

Typical vectors useful for expression of genes in higher plants are well known in the art and include vectors derived from the tumor-inducing (Ti) plasmid of *Agrobacterium tumefaciens* described (Rogers *et al.*, 1987). However, several other plant integrating vector systems are known to function in plants including pCaMVCN transfer control vector described (Fromm *et al.*, 1985). Plasmid pCaMVCN (available from Pharmacia, Piscataway, NJ) includes the cauliflower mosaic virus CaMV 35S promoter.

In preferred embodiments, the vector used to express the polypeptide includes a selection marker that is effective in a plant cell, preferably a drug resistance selection marker. One preferred drug resistance marker is the gene whose expression results in kanamycin resistance; *i.e.*, the chimeric gene containing the nopaline synthase promoter, Tn5 neomycin phosphotransferase II (*nptII*) and nopaline synthase 3′ non-translated region described (Rogers *et al.*, 1988). Another preferred selection marker is the bar gene, which confers glyphosphate resistance. Selectable markers useful in plant transformation are disclosed in U.S. Pat. Appl. No. 08/113,561, filed August 25, 1993, which is specifically incorporated herein by reference.

RNA polymerase transcribes a coding DNA sequence through a site where polyadenylation occurs. Typically, DNA sequences located a few hundred base pairs downstream of the polyadenylation site serve to terminate transcription. Those DNA sequences

are referred to herein as transcription-termination regions. Those regions are required for efficient polyadenylation of transcribed messenger RNA (mRNA).

Means for preparing expression vectors are well known in the art. Expression (transformation vectors) used to transform plants and methods of making those vectors are described in U. S. Patent Nos. 4,971,908, 4,940,835, 4,769,061 and 4,757,011, the disclosures of which are incorporated herein by reference. Those vectors can be modified to include a coding sequence in accordance with the present invention.

A variety of methods have been developed to operatively link DNA to vectors *via* complementary cohesive termini or blunt ends. For instance, complementary homopolymer tracts can be added to the DNA segment to be inserted and to the vector DNA. The vector and DNA segment are then joined by hydrogen bonding between the complementary homopolymeric tails to form recombinant DNA molecules.

A coding region that encodes a polypeptide having the ability to confer increased levels of lignin biosynthesis to a cell is preferably a tomato IEMT gene. In preferred embodiments, such a polypeptide has the amino acid residue sequence of SEQ ID NO:2 or a functional equivalent of that sequence.

In other embodiments a IEMT gene of the current invention may be operable linked to a transit peptide. In a particular embodiment, this transit peptide may be directed to the chloroplast. The transit peptide may, for example, be a heterologous transit peptide such as that of the small subunit of ribulose bisphosphate carboxylase. In still other embodiments of the current invention, a transit peptide operable linked to a IEMT gene may be directed to locations within a plant cell other than the flowers.

VIII. Transgenic IEMT Plants and Plant Cells

The present invention provides methods for producing a transgenic plant which comprises a nucleic acid segment encoding at least a portion of the IEMT gene of the present invention. The process of producing transgenic plants is well-known in the art. In general, the method comprises transforming a suitable host cell with a DNA segment which contains a promoter operatively linked to a coding region that encodes a IEMT gene. Such a coding region is generally operatively linked to a transcription-terminating region, whereby the promoter is capable of driving the transcription of the coding region in the cell, and hence providing the cell the ability to produce the recombinant protein *in vivo*. Alternatively, in instances where it is desirable to control, regulate, or decrease the amount of a particular recombinant protein expressed in a particular transgenic cell, the invention also provides for the expression of IEMT protein antisense mRNA. The use of antisense mRNA as a means of controlling or decreasing the amount of a given protein of interest in a cell is well-known in the art.

Another aspect of the invention comprises transgenic plants which express a gene or gene segment encoding one or more of the novel polypeptide compositions disclosed herein. As used herein, the term "transgenic plant" is intended to refer to a plant that has incorporated DNA sequences, which are in addition to those originally present, DNA sequences not normally transcribed into RNA or translated into a protein ("expressed"), or any other genes or DNA sequences which one desires to introduce into the non-transformed plant, such as genes which may normally be present in the non-transformed plant but which one desires to either genetically engineer or to have altered expression.

It is contemplated that in some instances the genome of a transgenic plant of the present invention will have been augmented through the stable introduction of one or more lignin biosynthesis genes, either native, synthetically modified, or mutated. In some instances, more than one transgene will be incorporated into the genome of the transformed host plant cell. Such is the case when more than one lignin biosynthesis protein-encoding DNA segment is

incorporated into the genome of such a plant. In certain situations, it may be desirable to have one, two, three, four, five or even more lignin biosynthesis proteins (either native or recombinantly-engineered) incorporated and stably expressed in the transformed transgenic plant.

A preferred gene which may be introduced includes, for example, the IEMT1 gene in order to reduce the clove-like taste and flavor in plants. Still other preferred genes are those which, when introduced into a plant or plant cell, result in increased methylation of isoeugenol and eugenol, which will alter the flavor of the plant. Still other preferred genes will be those genes that methylate ferulic acid, coniferyl aldehyde or coniferyl alcohols of the lignin biosynthetic pathway, resulting in a decrease in precursors available for lignin biosynthesis and increasing the precusors available for phenylpropanoid biosynthesis. These genes may be derived from wild-type IEMT, or may be chimeras of IEMT and COMT.

Means for transforming a plant cell and the preparation of a transgenic cell line are well-known in the art, and are discussed herein. Vectors, plasmids, cosmids, YACs (yeast artificial chromosomes) and DNA segments for use in transforming such cells will, of course, generally comprise either the operons, genes, or gene-derived sequences of the present invention, either native, or synthetically-derived, and particularly those encoding the disclosed crystal proteins. These DNA constructs can further include structures such as promoters, enhancers, polylinkers, or even gene sequences which have positively- or negatively-regulating activity upon the particular genes of interest as desired. The DNA segment or gene may encode either a native or modified crystal protein, which will be expressed in the resultant recombinant cells, and/or which will impart an improved phenotype to the regenerated plant.

Such transgenic plants may be desirable for their decreased clove-like flavor and fragrance, resulting from the incorporation of one or more DNA segments encoding one or more IEMT and other phenylpropanoid biosynthesis genes. Particularly preferred plants include tomato and other plants the like. These transgenic plants also will be desireable for reduced lignin content, resulting from the incorporation a gene that is capable of methylating, at

the 4' position, various lignin precursors such as *p*-coumaric acid, caffeic acid, ferulic acid, coniferyl aldehyde and coniferyl alcohol. Particularly preferred plants include those having a high lignin content, such as wheat, oilseed rape, sugar beet, maize, alfalfa, forage grasses and sunflower, and also tree crops such as eucalyptus, pine species and poplar, as well as other commercially useful plants used in the production of yarns and paper goods, known to those of skill in the art (see US Patent No. 5,451,514, incorporated herein by reference).

In a related aspect, the present invention also encompasses a seed produced by the transformed plant, a progeny from such seed, and seed produced by the progeny of the original transgenic plant, produced in accordance with the above process. Such progeny and seeds will have a lignin or phenylpropanoid biosynthesis protein-encoding transgene stably incorporated into their genome, and such progeny plants will inherit the traits afforded by the introduction of a stable transgene in Mendelian fashion. All such transgenic plants having incorporated into their genome transgenic DNA segments encoding one or more lignin or phenylpropanoid biosynthesis proteins or polypeptides are aspects of this invention.

A. Sources of Cells

Practicing the present invention includes the generation and use of recipient cells. As used herein, the term "recipient cells" refers to cells that are receptive to transformation and subsequent regeneration into stably transformed, fertile plants. Maize recipient cell targets include, but are not limited to, meristem cells, Type I, Type II, and Type III callus, immature embryos and gametic cells such as microspores pollen, sperm and egg cells. Type I, Type II, and Type III callus may be initiated from tissue sources including, but not limited to, immature embryos, seedling apical meristems, microspores and the such. Those cells which are capable of proliferating as callus are also recipient cells for genetic transformation. Pollen, as well as its precursor cells, microspores, may be capable of functioning as recipient cells for genetic transformation, or as vectors to carry foreign DNA for incorporation during fertilization. Direct pollen transformation would obviate the need for cell culture. Meristematic cells (*i.e.*, plant cells capable of continual cell division and characterized by an undifferentiated cytological appearance, normally found at growing points or tissues in plants such as root tips, stem apices,

lateral buds, *etc.*) may represent another type of recipient plant cell. Because of their undifferentiated growth and capacity for organ differentiation and totipotency, a single transformed meristematic cell could be recovered as a whole transformed plant. In fact, it is proposed that embryogenic suspension cultures may be an *in vitro* meristematic cell system, retaining an ability for continued cell division in an undifferentiated state, controlled by the media environment.

In certain embodiments, cultured plant cells that can serve as recipient cells for transforming with desired DNA segments include corn cells, and more specifically, cells from *Zea mays* L. Somatic cells are of various types. Embryogenic cells are one example of somatic cells which may be induced to regenerate a plant through embryo formation. Non-embryogenic cells are those which will typically not respond in such a fashion. An example of non-embryogenic cells are certain Black Mexican Sweet (BMS) corn cells.

The present invention also provides certain techniques that may enrich recipient cells within a cell population. For example, Type II callus development, followed by manual selection and culture of friable, embryogenic tissue, generally results in an enrichment of recipient cells for use in, *e.g.*, micro-projectile transformation. Suspension culturing, particularly using the media disclosed herein, may also improve the ratio of recipient to non-recipient cells in any given population. Manual selection techniques which are employed to select recipient cells may include, *e.g.*, assessing cell morphology and differentiation, or may use various physical or biological means. Cryopreservation is also contemplated as a possible method of selecting for recipient cells.

Manual selection of recipient cells, *e.g.*, by selecting embryogenic cells from the surface of a Type II callus, is one means employed by the inventors in an attempt to enrich for recipient cells prior to culturing (whether cultured on solid media or in suspension). The preferred cells may be those located at the surface of a cell cluster, and may further be identifiable by their lack of differentiation, their size and dense cytoplasm. The preferred cells will generally be those cells which are less differentiated, or not yet committed to differentiation. Thus, one may

wish to identify and select those cells which are cytoplasmically dense, relatively unvacuolated with a high nucleus to cytoplasm ratio (*e.g.*, determined by cytological observations), small in size (*e.g.*, 10-20 mm), and capable of sustained divisions and somatic proembryo formation.

It is proposed that other means for identifying such cells may also be employed. For example, through the use of dyes, such as Evan's blue, which are excluded by cells with relatively non-permeable membranes, such as embryogenic cells, and taken up by relatively differentiated cells such as root-like cells and snake cells (so-called due to their snake-like appearance).

Other possible means of identifying recipient cells include the use of isozyme markers of embryogenic cells, such as glutamate dehydrogenase, which can be detected by cytochemical stains (Fransz *et al.*, 1989). However, it is cautioned that the use of isozyme markers such as glutamate dehydrogenase may lead to some degree of false positives from non-embryogenic cells such as rooty cells which nonetheless have a relatively high metabolic activity.

B. Media

In certain embodiments, recipient cells are selected following growth in culture. Where employed, cultured cells will preferably be grown either on solid supports or in the form of liquid suspensions. In either instance, nutrients may be provided to the cells in the form of media, and environmental conditions controlled. There are many types of tissue culture media comprised of amino acids, salts, sugars, growth regulators and vitamins. Most of the media employed in the practice of the invention will have some similar components, the media differ in the composition and proportions of their ingredients depending on the particular application envisioned. For example, various cell types usually grow in more than one type of media, but will exhibit different growth rates and different morphologies, depending on the growth media. In some media, cells survive but do not divide.

Various types of media suitable for culture of plant cells have been previously described. Examples of these media include, but are not limited to, the N6 medium described

by Chu *et al.* (1975) and MS media (Murashige and Skoog, 1962). The inventors have discovered that media such as MS which have a high ammonia/nitrate ratio are counterproductive to the generation of recipient cells in that they promote loss of morphogenic capacity. N6 media, on the other hand, has a somewhat lower ammonia/nitrate ratio, and is contemplated to promote the generation of recipient cells by maintaining cells in a proembryonic state capable of sustained divisions.

C. Cell Cultures

1. Initiation

In the practice of the invention it is sometimes, but not always, necessary to develop cultures which contain recipient cells. Suitable cultures can be initiated from a number of whole plant tissue explants including, but not limited to, immature embryos, leaf bases, immature tassels, anthers, microspores, and other tissues containing cells capable of *in vitro* proliferation and regeneration of fertile plants. For example, recipient cell cultures are initiated from immature embryos by growing excised immature embryos on a solid culture medium containing growth regulators including, but not limited to, dicamba., 2,4-D, NAA, and IAA. In some instances it will be preferred to add silver nitrate to culture medium for callus initiation as this compound has been reported to enhance culture initiation (Vain et al., 1989). Embryos will produce callus that varies greatly in morphology including from highly unorganized cultures containing very early embryogenic structures (such as, but not limited to, type II cultures in maize), to highly organized cultures containing large late embryogenic structures (such as, but not limited to, type I cultures in maize). This variation in culture morphology may be related to genotype, culture medium composition, size of the initial embryos and other factors. Each of these types of culture morphologies is a source of recipient cells.

The development of suspension cultures capable of plant regeneration may be used in the context of the present invention. Suspension cultures may be initiated by transferring callus tissue to liquid culture medium containing growth regulators. Addition of coconut water or other substances to suspension culture medium may enhance growth and culture morphology, but the utility of suspension cultures is not limited to those containing these compounds. In

some embodiments of this invention, the use of suspension cultures will be preferred as these cultures grow more rapidly and are more easily manipulated than callus cells growing on solid culture medium.

When immature embryos or other tissues directly removed from a whole plant are used as the target tissue for DNA delivery, it will only be necessary to initiate cultures of cells insofar as is necessary for identification and isolation of transformants. In an illustrative embodiment, DNA is introduced by particle bombardment into immature embryos following their excision from the plant . Embryos are transferred to a culture medium that will support proliferation of tissues and allow for selection of transformed sectors, 0-14 days following DNA delivery. In this embodiment of the invention it is not necessary to establish stable callus cultures capable of long term maintenance and plant regeneration.

2. Maintenance

The method of maintenance of cell cultures may contribute to their utility as sources of recipient cells for transformation. Manual selection of cells for transfer to fresh culture medium, frequency of transfer to fresh culture medium, composition of culture medium, and environment factors including, but not limited to, light quality and quantity and temperature are all important factors in maintaining callus and/or suspension cultures that are useful as sources of recipient cells. It is contemplated that alternating callus between different culture conditions may be beneficial in enriching for recipient cells within a culture. For example, it is proposed that cells may be cultured in suspension culture, but transferred to solid medium at regular intervals. After a period of growth on solid medium cells can be manually selected for return to liquid culture medium. It is proposed that by repeating this sequence of transfers to fresh culture medium it is possible to enrich for recipient cells. It is also contemplated that passing cell cultures through a 1.9 mm sieve is useful in maintaining the friability of a callus or suspension culture and may be beneficial is enriching for transformable cells.

3. Cryopreservation

Additionally, the inventors propose that cryopreservation may effect the development of, or perhaps select for, recipient cells. Cryopreservation selection may operate due to a selection against highly vacuolated, non-embryogenic cells, which may be selectively killed during cryopreservation. The inventors propose that there is a temporal window in which cultured cells retain their regenerative ability, thus, it is believed that they must be preserved at or before that temporal period if they are to be used for future transformation and regeneration.

For use in transformation, suspension or callus culture cells may be cryopreserved and stored for periods of time, thawed, then used as recipient cells for transformation. An illustrative embodiment of cryopreservation methods comprises the steps of slowly adding cryoprotectants to suspension cultures to give a final concentration of 10% dimethyl sulfoxide, 10% polyethylene glycol (6000MW), O.23 M proline and 0.23 M glucose. The mixture is then cooled to -35°C at 0.5°C per min. After an isothermal period of 45 min, samples are placed in liquid N_2 (modification of methods of Finkle *et al.*(1985)). To reinitiate suspension cultures from cryopreserved material, cells may be thawed rapidly and pipetted onto feeder plates similar to those described by Rhodes *et al.* (1988).

IX. DNA Sequences

Virtually any DNA composition may be used for delivery to recipient cells by modern transformation techniques to ultimately produce fertile transgenic plants. In accordance with the present invention the genes used to create transgenic plants will be IEMT genes of phenylpropanoid biosynthesis. By way of example, DNA segments in the form of vectors and plasmids, or linear DNA fragments, in some instances containing only the IEMT gene DNA element to be expressed in the plant, and the like, may be employed.

In certain embodiments, it is contemplated that one may wish to employ replication-competent viral vectors in monocot transformation. Such vectors include, for example, wheat dwarf virus (WDV) "shuttle" vectors, such as pW1-11 and PW1-GUS (Ugaki *et al.*, 1991).

These vectors are capable of autonomous replication in maize cells as well as *E. coli*, and as such may provide increased sensitivity for detecting DNA delivered to transgenic cells. A replicating vector may also be useful for delivery of genes flanked by DNA sequences from transposable elements such as Ac, Ds, or Mu. It has been proposed (Laufs et al., 1990) that transposition of these elements within the maize genome requires DNA replication. It is also contemplated that transposable elements would be useful for introducing DNA fragments lacking elements necessary for selection and maintenance of the plasmid vector in bacteria, *e.g.*, antibiotic resistance genes and origins of DNA replication. It is also proposed that use of a transposable element such as Ac, Ds, or Mu would actively promote integration of the desired DNA and hence increase the frequency of stably transformed cells.

Vectors, plasmids, cosmids, YACs (yeast artificial chromosomes) and DNA segments for use in transforming such cells will, of course, generally comprise the cDNA, gene or genes which one desires to introduce into the cells. These DNA constructs can further include structures such as promoters, enhancers, polylinkers, or even regulatory genes as desired. The DNA segment or gene chosen for cellular introduction will encode a protein involved in phenylpropanoid biosynthesis and will be expressed in the resultant recombinant cells, such as will result in a screenable or selectable trait and/or which will impart an improved phenotype to the regenerated plant.

A. Regulatory Elements

The construction of vectors which may be employed in conjunction with the present invention will be known to those of skill of the art in light of the present disclosure (see *e.g.*, Sambrook *et al.*, 1989; Gelvin *et al.*, 1990). Preferred constructs will generally include a plant promoter such as the CaMV 35S promoter (Odell *et al.*, 1985), or others such as CaMV 19S (Lawton *et al.*, 1987), *nos* (Ebert *et al.*, 1987), Adh (Walker *et al.*, 1987), sucrose synthase (Yang and Russell, 1990), a-tubulin, actin (Wang et al., 1992), *cab* (Sullivan *et al.*, 1989), PEPCase (Hudspeth and Grula, 1989) or those associated with the R gene complex (Chandler *et al.*, 1989). Tissue specific promoters such as root cell promoters (Conkling *et al.*, 1990) and

tissue specific enhancers (Fromm *et al.*, 1989) are also contemplated to be particularly useful, as are inducible promoters such as ABA- and turgor-inducible prompters.

Constructs will also include the gene of interest along with a 3' end DNA sequence that acts as a signal to terminate transcription and allow for the poly-adenylation of the resultant mRNA. The most preferred 3' elements are contemplated to be those from the nopaline synthase gene of *Agrobacterium tumefasciens* (Bevan *et al.*, 1983), the terminator for the T7 transcript from the octopine synthase gene of *Agrobacterium tumefasciens*, and the 3' end of the protease inhibitor I or II genes from potato or tomato. Regulatory elements such as Adh intron 1 (Callis *et al.*, 1987), sucrose synthase intron (Vasil *et al.*, 1989) or TMV omega element (Gallie, *et al.*, 1989), or actin intron (Wang et al., 1992) ,may further be included where desired.

As the DNA sequence between the transcription initiation site and the start of the coding sequence, *i.e.*, the untranslated leader sequence, can influence gene expression, one may also wish to employ a particular leader sequence. Preferred leader sequences are contemplated to include those which include sequences predicted to direct optimum expression of the attached gene, *i.e.*, to include a preferred consensus leader sequence which may increase or maintain mRNA stability and prevent inappropriate initiation of translation. The choice of such sequences will be known to those of skill in the art in light of the present disclosure. Sequences that are derived from genes that are highly expressed in plants, and in maize in particular, will be most preferred.

It is contemplated that vectors for use in accordance with the present invention may be constructed to include the *ocs* enhancer element. This element was first identified as a 16 bp palindromic enhancer from the octopine synthase (*ocs*) gene of agrobacterium (Ellis *et al.*, 1987), and is present in at least 10 other promoters (Bouchez *et al.*, 1989). It is proposed that the use of an enhancer element, such as the *ocs* element and particularly multiple copies of the element, will act to increase the level of transcription from adjacent promoters when applied in the context of monocot transformation.

It is specifically envisioned that IEMT genes may be introduced under the control of novel promoters or enhancers, *etc.*, or perhaps even homologous or tissue specific (*e.g.*, root-, collar/sheath-, whorl-, stalk-, earshank-, kernel- or leaf-specific) promoters or control elements. Indeed, it is envisioned that a particular use of the present invention will be the targeting of lignin or phenylpropanoid biosynthesis in a tissue-specific manner. For example, insect resistant genes may be expressed specifically in the whorl and collar/sheath tissues which are targets for the first and second broods, respectively, of ECB.

Vectors for use in tissue-specific targeting of genes in transgenic plants will typically include tissue-specific promoters and may also include other tissue-specific control elements such as enhancer sequences. Promoters which direct specific or enhanced expression in certain plant tissues will be known to those of skill in the art in light of the present disclosure. These include, for example, the *rbcS* promoter, specific for green tissue; the *ocs*, *nos* and *mas* promoters which have higher activity in roots or wounded leaf tissue; a truncated (-90 to +8) 35S promoter which directs enhanced expression in roots, an a-tubulin gene that directs expression in roots and promoters derived from zein storage protein genes which direct expression in endosperm. It is particularly contemplated that one may advantageously use the 16 bp *ocs* enhancer element from the octopine synthase (*ocs*) gene (Ellis *et al.*, 1987; Bonchez *et al.*, 1989), especially when present in multiple copies, to achieve enhanced expression in roots.

It is also contemplated that tissue specific expression may be functionally accomplished by introducing a constitutively expressed gene (all tissues) in combination with an antisense gene that is expressed only in those tissues where the gene product is not desired. For example, a gene coding for a IEMT gene may be introduced such that it is expressed in all tissues using the 35S promoter from Cauliflower Mosaic Virus. Expression of an antisense transcript of the same IEMT gene in a tomato plant, using for example a fruit-specific promoter, would prevent accumulation of eugenol and methyl isoeugenol in the fruit thereby eliminating the clove-like flavor.

Alternatively, one may wish to obtain novel tissue-specific promoter sequences for use in accordance with the present invention. To achieve this, one may first isolate cDNA clones from the tissue concerned and identify those clones which are expressed specifically in that tissue, for example, using Northern blotting. Ideally, one would like to identify a gene that is not present in a high copy number, but which gene product is relatively abundant in specific tissues. The promoter and control elements of corresponding genomic clones may then be localized using the techniques of molecular biology known to those of skill in the art.

It is contemplated that expression of IEMT genes in transgenic plants may in some cases be desired only under specified conditions. It is contemplated that expression of such genes at high levels may have detrimental effects. It is known that a large number of genes exist that respond to the environment. For example, expression of some genes such as *rbcS*, encoding the small subunit of ribulose bisphosphate carboxylase, is regulated by light as mediated through phytochrome. Other genes are induced by secondary stimuli. A number of genes have been shown to be induced by ABA (Skriver and Mundy, 1990). It is also expected, in particular embodiments, that inducible expression of IEMT genes in transgenic plants may be desired.

It is proposed that in some embodiments of the present invention expression of a IEMT gene in a transgenic plant will be desired only in a certain time period during the development of the plant. Developmental timing is frequently correlated with tissue specific gene expression. For example, expression of zein storage proteins is initiated in the endosperm about 15 days after pollination.

Additionally, vectors may be constructed and employed in the intracellular targeting of a specific gene product within the cells of a transgenic plant or in directing a protein to the extracellular environment. This will generally be achieved by joining a DNA sequence encoding a transit or signal peptide sequence to the coding sequence of a particular gene. The resultant transit, or signal, peptide will transport the protein to a particular intracellular, or

extracellular destination, respectively, and will then be post-translationally removed. Transit or signal peptides act by facilitating the transport of proteins through intracellular membranes, *e.g.*, vacuole, vesicle, plastid and mitochondrial membranes, whereas signal peptides direct proteins through the extracellular membrane.

A particular example of such a use concerns the direction of a herbicide resistance gene, such as the EPSPS gene, to a particular organelle such as the chloroplast rather than to the cytoplasm. This is exemplified by the use of the *rbcS* transit peptide which confers plastid-specific targeting of proteins. In addition, it is proposed that it may be desirable to target IEMT genes to the extracellular spaces or to the vacuole.

It is also contemplated that it may be useful to target DNA itself within a cell. For example, it may be useful to target introduced DNA to the nucleus as this may increase the frequency of transformation. Within the nucleus itself it would be useful to target a gene in order to achieve site specific integration. For example, it would be useful to have an gene introduced through transformation replace an existing gene in the cell.

B. Marker Genes

In order to improve the ability to identify transformants, one may desire to employ a selectable or screenable marker gene as, or in addition to, the expressible gene of interest. "Marker genes" are genes that impart a distinct phenotype to cells expressing the marker gene and thus allow such transformed cells to be distinguished from cells that do not have the marker. Such genes may encode either a selectable or screenable marker, depending on whether the marker confers a trait which one can select for by chemical means, *i.e.*, through the use of a selective agent (*e.g.*, a herbicide, antibiotic, or the like), or whether it is simply a trait that one can identify through observation or testing, *i.e.*, by screening (*e.g.*, the R-locus trait). Of course, many examples of suitable marker genes are known to the art and can be employed in the practice of the invention.

Included within the terms selectable or screenable marker genes are also genes which encode a "secretable marker" whose secretion can be detected as a means of identifying or selecting for transformed cells. Examples include markers which encode a secretable antigen that can be identified by antibody interaction, or even secretable enzymes which can be detected by their catalytic activity. Secretable proteins fall into a number of classes, including small, diffusible proteins detectable, *e.g.*, by ELISA; small active enzymes detectable in extracellular solution (*e.g.*, α-amylase, β-lactamase, phosphinothricin acetyltransferase); and proteins that are inserted or trapped in the cell wall (*e.g.*, proteins that include a leader sequence such as that found in the expression unit of extensin or tobacco PR-S).

With regard to selectable secretable markers, the use of a gene that encodes a protein that becomes sequestered in the cell wall, and which protein includes a unique epitope is considered to be particularly advantageous. Such a secreted antigen marker would ideally employ an epitope sequence that would provide low background in plant tissue, a promoter-leader sequence that would impart efficient expression and targeting across the plasma membrane, and would produce protein that is bound in the cell wall and yet accessible to antibodies. A normally secreted wall protein modified to include a unique epitope would satisfy all such requirements.

One example of a protein suitable for modification in this manner is extensin, or hydroxyproline rich glycoprotein (HPRG). The use of the maize HPRG (Steifel *et al.*, 1990) which is preferred as this molecule is well characterized in terms of molecular biology, expression and protein structure. However, any one of a variety of extensions and/or glycine-rich wall proteins (Keller *et al.*, 1989) could be modified by the addition of an antigenic site to create a screenable marker.

One exemplary embodiment of a secretable screenable marker concerns the use of the maize genomic clone encoding the wall protein HPRG, modified to include the unique 15 residue epitope MATVPELNCEMPPSD (SEQ ID NO:7) from the pro-region of murine interleukin-1-ß (IL-1-ß). However, virtually any detectable epitope may be employed in such

embodiments, as selected from the extremely wide variety of antigen:antibody combinations known to those of skill in the art. The unique extracellular epitope, whether derived from IL-1-ß or any other protein or epitopic substance, can then be straightforwardly detected using antibody labeling in conjunction with chromogenic or fluorescent adjuncts.

Elements of the present disclosure are exemplified in detail through the use of the *bar* and/or GUS genes, and also through the use of various other markers. Of course, in light of this disclosure, numerous other possible selectable and/or screenable marker genes will be apparent to those of skill in the art in addition to the one set forth hereinbelow. Therefore, it will be understood that the following discussion is exemplary rather than exhaustive. In light of the techniques disclosed herein and the general recombinant techniques which are known in the art, the present invention renders possible the introduction of any gene, including marker genes, into a recipient cell to generate a transformed plant.

C. Selectable Markers

Possible selectable markers for use in connection with the present invention include, but are not limited to, a *neo* gene (Potrykus *et al.*, 1985) which codes for kanamycin resistance and can be selected for using kanamycin, G418, *etc.*; a *bar* gene which codes for bialaphos resistance; a mutant gene which encodes an altered EPSP synthase protein (Hinchee *et al.*, 1988) thus conferring glyphosate resistance; a nitrilase gene such as *bxn* from *Klebsiella ozaenae* which confers resistance to bromoxynil (Stalker *et al.*, 1988); a mutant acetolactate synthase gene (ALS) which confers resistance to imidazolinone, sulfonylurea or other ALS inhibiting chemicals (European Patent Application 154,204, 1985); a methotrexate resistant DHFR gene (Thillet *et al.*, 1988), or a dalapon dehalogenase gene that confers resistance to the herbicide dalapon; or a mutated anthranilate synthase gene that confers resistance to 5-methyl tryptophan. Where a mutant EPSP synthase gene is employed, additional benefit may be realized through the incorporation of a suitable chloroplast transit peptide, CTP (European Pat. No. 0189707).

An illustrative embodiment of a selectable marker gene capable of being used in systems to select transformants is the genes that encode the enzyme phosphinothricin acetyltransferase, such as the *bar* gene from *Streptomyces hygroscopicus* or the *pat* gene from *Streptomyces viridochromogenes*. The enzyme phosphinothricin acetyl transferase (PAT) inactivates the active ingredient in the herbicide bialaphos, phosphinothricin (PPT). PPT inhibits glutamine synthetase, (Murakami *et al.*, 1986; Twell *et al.*, 1989) causing rapid accumulation of ammonia and cell death.

Where one desires to employ a bialaphos resistance gene in the practice of the invention, the inventors have discovered that a particularly useful gene for this purpose is the *bar* or *pat* genes obtainable from species of Streptomyces (*e.g.*, ATCC No. 21,705). The cloning of the *bar* gene has been described (Murakami *et al.*, 1986; Thompson *et al.*, 1987) as has the use of the *bar* gene in the context of plants other than monocots (De Block *et al.*, 1987; De Block *et al.*, 1989).

D. Screenable Markers

Screenable markers that may be employed include a β-glucuronidase or *uid*A gene (GUS) which encodes an enzyme for which various chromogenic substrates are known; an R-locus gene, which encodes a product that regulates the production of anthocyanin pigments (red color) in plant tissues (Dellaporta *et al.*, 1988); a β-lactamase gene (Sutcliffe, 1978), which encodes an enzyme for which various chromogenic substrates are known (*e.g.*, PADAC, a chromogenic cephalosporin); a *xyl*E gene (Zukowsky *et al.*, 1983) which encodes a catechol dioxygenase that can convert chromogenic catechols; an α-amylase gene (Ikuta *et al.*, 1990); a tyrosinase gene (Katz *et al.*, 1983) which encodes an enzyme capable of oxidizing tyrosine to DOPA and dopaquinone which in turn condenses to form the easily-detectable compound melanin; a β-galactosidase gene, which encodes an enzyme for which there are chromogenic substrates; a luciferase (*lux*) gene (Ow *et al.*, 1986), which allows for bioluminescence detection; a green fluorescent protein (GFP) gene, or even an aequorin gene (Prasher *et al.*, 1985), which may be employed in calcium-sensitive bioluminescence detection.

A screenable marker contemplated for use in the present invention is firefly luciferase, encoded by the *lux* gene. The presence of the *lux* gene in transformed cells may be detected using, for example, X-ray film, scintillation counting, fluorescent spectrophotometry, low-light video cameras, photon counting cameras or multiwell luminometry. It is also envisioned that this system may be developed for populational screening for bioluminescence, such as on tissue culture plates, or even for whole plant screening.

It is further contemplated that a gene encoding green fluorescent protein (GFP) could also be used as a screenable marker. Cells expressing GFP fluoresce when illuminated with light of particular wavelengths, especially ultraviolet light. Cells or plants expressing GFP can thereby be readily identified..

E. Negative Selectable Markers

It is contemplated that in particular embodiments a negative selectable marker may be used with the current invention. It is contemplated that when two or more genes are introduced together by co-transformation that the genes will be linked together on the host chromosome. For example, a gene encoding a IEMT protein that confers a clove-like flavor on the plant may be introduced into a plant together with a *bar* gene that is useful as a selectable marker and confers resistance to the herbicide Ignite® on the plant. However, it may not be desirable to have an insect resistant plant that is also resistant to the herbicide Ignite®. It is proposed that one could also introduce an antisense *bar* gene that is expressed in those tissues where one does not want expression of the *bar* gene, *e.g.*, in whole plant parts. Hence, although the *bar* gene is expressed and is useful as a selectable marker, it is not useful to confer herbicide resistance on the whole plant. The *bar* antisense gene is a negative selectable marker.

It is also contemplated that a negative selection is necessary in order to screen a population of transformants for rare homologous recombinants generated through gene targeting. For example, a homologous recombinant may be identified through the inactivation of a gene that was previously expressed in that cell. The antisense gene to neomycin phosphotransferase II (nptII) has been investigated as a negative selectable marker in tobacco

(*Nicotiana tabacum*) and *Arabidopsis thaliana* (Xiang, C. and Guerra, D.J. 1993). In this example both sense and antisense npt II genes are introduced into a plant through transformation and the resultant plants are sensitive to the antibiotic kanamycin. An introduced gene that integrates into the host cell chromosome at the site of the antisense nptII gene, and inactivates the antisense gene, will make the plant resistant to kanamycin and other aminoglycoside antibiotics. Therefore, rare site specific recombinants may be identified by screening for antibiotic resistance. Similarly, any gene, native to the plant or introduced through transformation, that when inactivated confers resistance to a compound, may be useful as a negative selectable marker.

X. Transformation Techniques

A. Electroporation

The application of brief, high-voltage electric pulses to a variety of bacterial, animal and plant cells leads to the formation of nanometer-sized pores in the plasma membrane. DNA is taken directly into the cell cytoplasm either through these pores or as a consequence of the redistribution of membrane components that accompanies closure of the pores. Electroporation can be extremely efficient and can be used both for transient expression of cloned genes and for establishment of cell lines that carry integrated copies of the gene of interest. Electroporation, in contrast to calcium phosphate-mediated transfection and protoplast fusion, frequently gives rise to cell lines that carry one, or at most a few, integrated copies of the foreign DNA.

The introduction of DNA by means of electroporation, is well-known to those of skill in the art. In this method, certain cell wall-degrading enzymes, such as pectin-degrading enzymes, are employed to render the target recipient cells more susceptible to transformation by electroporation than untreated cells. Alternatively, recipient cells are made more susceptible to transformation, by mechanical wounding. To effect transformation by electroporation one may employ either friable tissues such as a suspension culture of cells, or embryogenic callus, or alternatively, one may transform immature embryos or other organized tissues directly. One would partially degrade the cell walls of the chosen cells by exposing them to pectin-degrading

enzymes (pectolyases) or mechanically wounding in a controlled manner. Such cells would then be recipient to DNA transfer by electroporation, which may be carried out at this stage, and transformed cells then identified by a suitable selection or screening protocol dependent on the nature of the newly incorporated DNA.

B. Microprojectile Bombardment

A further advantageous method for delivering transforming DNA segments to plant cells is microprojectile bombardment. In this method, particles may be coated with nucleic acids and delivered into cells by a propelling force. Exemplary particles include those comprised of tungsten, gold, platinum and the like.

An advantage of microprojectile bombardment, in addition to it being an effective means of reproducibly stably transforming monocots, is that neither the isolation of protoplasts (Christou *et al.*, 1988) nor the susceptibility to *Agrobacterium* infection is required. An illustrative embodiment of a method for delivering DNA into maize cells by acceleration is a Biolistics Particle Delivery System, which can be used to propel particles coated with DNA or cells through a screen, such as a stainless steel or Nytex screen, onto a filter surface covered with corn cells cultured in suspension. The screen disperses the particles so that they are not delivered to the recipient cells in large aggregates. It is believed that a screen intervening between the projectile apparatus and the cells to be bombarded reduces the size of projectiles aggregate and may contribute to a higher frequency of transformation by reducing damage inflicted on the recipient cells by projectiles that are too large.

For the bombardment, cells in suspension are preferably concentrated on filters or solid culture medium. Alternatively, immature embryos or other target cells may be arranged on solid culture medium. The cells to be bombarded are positioned at an appropriate distance below the macroprojectile stopping plate. If desired, one or more screens are also positioned between the acceleration device and the cells to be bombarded. Through the use of techniques set forth herein one may obtain up to 1000 or more foci of cells transiently expressing a marker

gene. The number of cells in a focus which express the exogenous gene product 48 hr post-bombardment often range from 1 to 10 and average 1 to 3.

In bombardment transformation, one may optimize the prebombardment culturing conditions and the bombardment parameters to yield the maximum numbers of stable transformants. Both the physical and biological parameters for bombardment are important in this technology. Physical factors are those that involve manipulating the DNA/microprojectile precipitate or those that affect the flight and velocity of either the macro- or microprojectiles. Biological factors include all steps involved in manipulation of cells before and immediately after bombardment, the osmotic adjustment of target cells to help alleviate the trauma associated with bombardment, and also the nature of the transforming DNA, such as linearized DNA or intact supercoiled plasmids. It is believed that pre-bombardment manipulations are especially important for successful transformation of immature embryos.

Accordingly, it is contemplated that one may wish to adjust various of the bombardment parameters in small scale studies to fully optimize the conditions. One may particularly wish to adjust physical parameters such as gap distance, flight distance, tissue distance, and helium pressure. One may also minimize the trauma reduction factors (TRFs) by modifying conditions which influence the physiological state of the recipient cells and which may therefore influence transformation and integration efficiencies. For example, the osmotic state, tissue hydration and the subculture stage or cell cycle of the recipient cells may be adjusted for optimum transformation. The execution of other routine adjustments will be known to those of skill in the art in light of the present disclosure.

C. Agrobacterium-Mediated Transfer

Agrobacterium-mediated transfer is a widely applicable system for introducing genes into plant cells because the DNA can be introduced into whole plant tissues, thereby bypassing the need for regeneration of an intact plant from a protoplast. The use of *Agrobacterium*-mediated plant integrating vectors to introduce DNA into plant cells is well known in the art. See, for example, the methods described (Fraley *et al.*, 1985; Rogers *et al.*, 1987). Further, the

integration of the Ti-DNA is a relatively precise process resulting in few rearrangements. The region of DNA to be transferred is defined by the border sequences, and intervening DNA is usually inserted into the plant genome as described (Spielmann *et al.*, 1986; Jorgensen *et al.*, 1987).

Modern *Agrobacterium* transformation vectors are capable of replication in *E. coli* as well as *Agrobacterium*, allowing for convenient manipulations as described in Klee *et al.*, 1985. Moreover, recent technological advances in vectors for *Agrobacterium*-mediated gene transfer have improved the arrangement of genes and restriction sites in the vectors to facilitate construction of vectors capable of expressing various polypeptide coding genes. The vectors described (Rogers *et al.*, 1987), have convenient multi-linker regions flanked by a promoter and a polyadenylation site for direct expression of inserted polypeptide coding genes and are suitable for present purposes. In addition, *Agrobacterium* containing both armed and disarmed Ti genes can be used for the transformations. In those plant strains where *Agrobacterium*-mediated transformation is efficient, it is the method of choice because of the facile and defined nature of the gene transfer.

Agrobacterium-mediated transformation of leaf disks and other tissues such as cotyledons and hypocotyls appears to be limited to plants that *Agrobacterium* naturally infects. *Agrobacterium*-mediated transformation is most efficient in dicotyledonous plants. Few monocots appear to be natural hosts for *Agrobacterium*, although transgenic plants have been produced in asparagus using *Agrobacterium* vectors as described (Bytebier *et al.*, 1987). Therefore, commercially important cereal grains such as rice, corn, and wheat must usually be transformed using alternative methods. Agrobacterium-mediated transformation of maize has, however, recently been described in U.S. Pat. No. 5,591,616, which is specifically incorporated herein by reference. Further, it has been shown that homozygous IEMT mutants are more susceptible to *Agrobacterium*-mediated transformation.

A transgenic plant formed using *Agrobacterium* transformation methods typically contains a single transgene or a few copies of a transgene on one chromosome. Such transgenic

plants can be referred to as being heterozygous for the added gene. However, inasmuch as use of the word "heterozygous" usually implies the presence of a complementary gene at the same locus of the second chromosome of a pair of chromosomes, and there is no such gene in a plant containing one added gene as here, it is believed that a more accurate name for such a plant is hemizygous.

More preferred is a transgenic plant that is homozygous for the added structural gene; *i.e.*, a transgenic plant that contains two added genes, one gene at the same locus on each chromosome of a chromosome pair. A homozygous transgenic plant can be obtained by sexually mating (selfing) a hemizygous transgenic plant that contains a single added gene, germinating some of the seed produced and analyzing the resulting plants produced for enhanced carboxylase activity relative to a control (native, non-transgenic) or a hemizygous transgenic plant.

It is to be understood that two different transgenic plants can also be mated to produce offspring that contain multiple independently segregating added, exogenous genes. Specifically contemplated by the inventors, is the creation of plants which contain 1,2,3,4,5, or even more independently segregating added, exogenous genes. Selfing of appropriate progeny can produce plants that are homozygous for all added, exogenous genes that encode a polypeptide of interest. Back-crossing to a parental plant and out-crossing with a non-transgenic plant are also contemplated.

D. Other Transformation Methods

Transformation of plant protoplasts can be achieved using methods based on calcium phosphate precipitation, polyethylene glycol treatment, electroporation, and combinations of these treatments (see, *e.g.*, Potrykus *et al.*, 1985; Lorz *et al.*, 1985; Fromm *et al.*, 1986; Uchimiya *et al.*, 1986; Callis *et al.*, 1987; Marcotte *et al.*, 1988).

Application of these systems to different plant strains depends upon the ability to regenerate that particular plant strain from protoplasts. Illustrative methods for the regeneration

of cereals from protoplasts are described (Fujimura *et al.*, 1985; Toriyama *et al.*, 1986; Yamada *et al.*, 1986; Abdullah *et al.*, 1986).

To transform plant strains that cannot be successfully regenerated from protoplasts, other ways to introduce DNA into intact cells or tissues can be utilized. For example, regeneration of cereals from immature embryos or explants can be effected as described (Vasil, 1988). In addition, "particle gun" or high-velocity microprojectile technology can be utilized (Vasil, 1992).

Using that latter technology, DNA is carried through the cell wall and into the cytoplasm on the surface of small metal particles as described (Klein *et al.*, 1987; McCabe *et al.*, 1988). The metal particles penetrate through several layers of cells and thus allow the transformation of cells within tissue explants.

XI. Methods for Producing Plants With Enhanced Fragrance and Flavor

By transforming a suitable host cell, such as a plant cell, with a recombinant IEMT gene-containing segment, expression of the encoded IEMT gene (*i.e.*, a protein catalyzing (methyl(iso)eugenol biosynthesis) can result in increased production of methylated forms of isoeugenol and eugenol, thereby reducing the clove-like flavor thereof. Additionally, IEMT expression or overexpression may facilitate the increased production of other desireable products such as vanillin.

It is also contemplated by the inventors that particular combinations of transformed phenylpropanoid biosynthesis genes may be created by standard plant breeding methods, which are well known in the art. Such breeding protocols may be aided by the use of genetic markers which are closely linked to the genes of interest.

A preferred method for transformation with IEMT genes envisioned by the inventors is microprojectile bombardment. Techniques for microprojectile bombardment are well known in

the art, and are described in Lundquist *et al.*, U.S. Pat. No. 5,538,880 and Adams *et al.* U.S. Pat. No. 5,489,520, which are specifically incorporated herein by reference.

By way of example, one may utilize an expression vector containing a coding region for a IEMT gene and an appropriate selectable marker to transform a suspension of embryogenic plant cells, such as wheat or corn cells using a method such as particle bombardment (Maddock *et al.*, 1991; Vasil *et al.*, 1992) to deliver the DNA coated on microprojectiles into the recipient cells. Transgenic plants are then regenerated from transformed embryogenic calli that express the proteins.

The formation of transgenic plants may also be accomplished using other methods of cell transformation which are known in the art such as *Agrobacterium*-mediated DNA transfer (Fraley *et al.*, 1983). Alternatively, DNA can be introduced into plants by direct DNA transfer into pollen (Zhou *et al.*, 1983; Hess, 1987; Luo *et al.*, 1988), by injection of the DNA into reproductive organs of a plant (Pena *et al.*, 1987), or by direct injection of DNA into the cells of immature embryos followed by the rehydration of desiccated embryos (Neuhaus *et al.*, 1987; Benbrook *et al.*, 1986).

The regeneration, development, and cultivation of plants from single plant protoplast transformants or from various transformed explants is well known in the art (Weissbach *et al.*, 1988). This regeneration and growth process typically includes the steps of selection of transformed cells, culturing those individualized cells through the usual stages of embryonic development through the rooted plantlet stage. Transgenic embryos and seeds are similarly regenerated. The resulting transgenic rooted shoots are thereafter planted in an appropriate plant growth medium such as soil. Protoplast methods of transformation of maize are described in U.S. Patent 5,350,689.

The development or regeneration of plants containing the foreign, exogenous gene that encodes a polypeptide of interest introduced by *Agrobacterium* from leaf explants can be achieved by methods well known in the art such as described in Horsch *et al.*, 1985. In this

procedure, transformants are cultured in the presence of a selection agent and in a medium that induces the regeneration of shoots in the plant strain being transformed as described (Fraley *et al.*, 1983).

This procedure typically produces shoots within two to four months and those shoots are then transferred to an appropriate root-inducing medium containing the selective agent and an antibiotic to prevent bacterial growth. Shoots that rooted in the presence of the selective agent to form plantlets are then transplanted to soil or other media to allow the production of roots. These procedures vary depending upon the particular plant strain employed, such variations being well known in the art.

Preferably, the regenerated plants are self-pollinated to provide homozygous transgenic plants, as discussed before. Otherwise, pollen obtained from the regenerated plants is crossed to seed-grown plants of agronomically important, preferably inbred lines. Conversely, pollen from plants of those important lines is used to pollinate regenerated plants. A transgenic plant of the present invention containing a desired polypeptide is cultivated using methods well known to one skilled in the art.

A transgenic plant of this invention thus has an increased amount of the IEMT protein, or the IEMT protein in combination with other lignin biosynthesis proteins, and thereby, an enhanced flux of lignin precursors such as ferulic acid, coniferyl aldehyde, coniferyl alcohol, methyleugenol and methylisoeugenol going through the pathway depicted in FIG. 3. A preferred transgenic plant is homozygous and can transmit particular genes and their activities to its progeny. A more preferred transgenic plant is homozygous for the foreign gene or genes, and transmits the gene or genes to all of its offspring on sexual mating. Seed from a transgenic plant may be grown in the field or greenhouse, and resulting sexually mature transgenic plants are self-pollinated to generate true breeding plants. The progeny from these plants become true breeding lines that are evaluated for, by way of example, decreased lignification and increased flavor and fragrance, preferably in the field, under a range of environmental conditions. The inventors contemplate that the present invention will find particular utility in the creation of transgenic plants of commercial interest including various turf grasses, wheat, corn, rice, barley,

oats, a variety of ornamental plants and vegetables, as well as a number of nut- and fruit-bearing trees and plants.

XIII. Examples

The following examples are included to demonstrate preferred embodiments of the invention. It should be appreciated by those of skill in the art that the techniques disclosed in the examples which follow represent techniques discovered by the inventor to function well in the practice of the invention, and thus can be considered to constitute preferred modes for its practice. However, those of skill in the art should, in light of the present disclosure, appreciate that many changes can be made in the specific embodiments which are disclosed and still obtain a like or similar result without departing from the spirit and scope of the invention.

EXAMPLE 1

Materials And Methods

Plant Material, Growth Conditions, Headspace Collection, and GC-MS Analysis

Details of the construction of true-breeding *Clarkia breweri* stocks, growing conditions, dynamic headspace collection on Tenax (Alltech, Inc., Deerfield, IL) and activated charcoal sorbents, and chemical analyses *via* GC-MS were as described by Raguso and Pichersky (1995). All headspace collections were performed in a growth chamber (Conviron, Asheville, NC) under a 12-h light/12-h dark photoperiod. Temperature was set to 22°C during the light period and 18°C during the dark period. In all studies, headspace collections from ambient air and from vegetative tissues were used as controls.

Time Course of Phenylpropanoid Production

Volatile phenylpropanoid production in individual flowers of four separate plants was monitored over a 6 day period beginning on the day before anthesis and continuing until floral abscission. Headspace volatiles were collected as described by Raguso and Pichersky (1995). The collections were made at 12-h intervals, corresponding to the dark and light periods in the growth chamber.

Localization and Quantitation of Phenylpropanoid Emission in Floral Parts

The specific floral parts responsible for scent emission were determined and the emission levels were quantified by headspace collection essentially as described by Raguso and Pichersky (1995). Headspace collection was made from attached, second-day (hermaphroditic) intact flowers and from same-stage flowers in which floral organs had been systematically removed, to leave only petals, only anthers, or only the pistil. To detect all volatiles emitted by a given flower part that could possibly emit different compounds at different times, a 24-h collection period was used.

IEMT Enzyme Extraction and Assay

(i) Enzyme Extraction

A crude protein extract was prepared by macerating flower parts in a microcentrifuge tube in the presence of ice-cold buffer (10 volumes fresh weight) containing 50 mM BisTris-HCl, pH 6.9, 10 mM 2-mercaptoethanol, 5 mM $Na_2S_2O_5$, 1% (w/v) PVP-40, and 10% (v/v) glycerol. The slurry was centrifuged for 10 min and the supernatant was transferred to a new tube. For each time point, flowers from three different plants were combined.

OMT Enzyme Assays and Product Analysis

Assay samples were prepared by adding to a 1.5-mL microcentrifuge tube: 10 µL of crude extract, 10 µL of assay buffer (250 mM Tris-HCl, pH 7.5, 25 mM $MgCl_2$, 10 mM DTT, 1 µL of 50 mM substrate [eugenol, isoeugenol, or other related compounds] in ethanol, 1.0 µL of *S*-[methyl-^{14}C]adenosyl-L-Met [40-60 mCi/mmol] in 10 mM sulfuric acid:ethanol; purchased from NEN Research Products, Boston, MA), and 28 µL of water to bring the assay volume to 50 µL. Assay samples were incubated at 30°C for 30 min in a heating block, after which 2.5 µL of 6 M HCl was added to stop the reaction. The radioactively labeled methylated product was extracted by the addition of 100 µL of ethyl acetate, and 20 µL of the organic phase (on top and clear in color) was transferred to a scintillation vial with 2 mL of nonaqueous scintillation fluid (Bio-Safe NA, Research Products International, Mount Prospect, IL) and counted in a liquid scintillation counter (model 2S6800, Beckman, Somerset, NJ]. The raw data (counts per

min) were converted to femtomoles of product produced per second based on the specific activity of the substrate, using the appropriate correction factors for counting efficiency.

To verify the identity of the products, organic extracts were analyzed in two ways. First, 20 μL was spotted on a 10 cm × 20 cm silica gel 60 F_{254} precoated TLC plate (EM Industries, Inc., Gibbstown, NJ), and 5 μL of a 5% (v/v) solution containing authentic methyleugenol or methylisoeugenol was spotted on the same plate as a standard. The plate was developed in a solvent system of 2:3:1 (v/v) benzene:acetic acid:water (De Carolis and Ibrahim, 1989). When the solvent was within 2 cm of the top edge of the plate, the plate was removed and allowed to dry. UV light revealed the elution points of the standards, from which retention-factor values were calculated.

TLC plates containing radioactively labeled product were analyzed by an imaging scanner (System 200, Bioscan, Inc., Washington, DC). Plates were scanned horizontally in lanes 0.6 cm in width for 5 min per lane by the Autochanger 3000 detector under a steady flow of 90% argon/10% methane at a flow rate of 70 kPa (10 p.s.i.). Bioscan computer software produced full-color, two-dimensional images of TLC plates using a color scale to display the location of radioactivity, indicating the presence of radioactively labeled product. Comparison of the retention-factor values of radioactive spots and those of the nonradioactive standards tentatively determined the identities of radioactively labeled enzyme product.

In addition, "cold assays" with nonradioactive SAM were performed by scaling up the reaction to a total volume of 1000 μL and a final substrate concentration of 1 mM. The products were organically extracted and analyzed by GC-MS analysis.

Isolation and Purification of OMTs from C. breweri

Preparation of crude extract from *C. breweri* petals was performed as previously described (Pichersky *et al.*, 1995). The enzyme activity was purified through successive chromatographic steps involving DE53 anion-exchange and hydroxyapatite columns as previously described (Pichersky *et al.*, 1995). The last step of purification utilized a 5'-ADP

affinity column from which the pyrophosphate group was removed with alkaline phosphatase as described by Attieh *et al.* (1995).

(i) IEMT protein purification

All manipulations were carried out at 4°C unless stated otherwise. In a typical purification procedure, 125 ml of crude extract was loaded onto a DEAE-cellulose column (10 ml of DE53, Whatman, Fairfield, NJ) that was pre-equilibrated with a solution containing 50 mM Bis-Tris, pH 6.9, 10% glycerol and 10 mM β-mercaptoethanol (buffer A) at a flow rate of 0.5 ml/min. After washing off unabsorbed material from the column with 30 ml of buffer A containing 100 mM KC1, IEMT was eluted with a step gradient of buffer A containing 0.15, 0.2, 0.25 and 0.3 M KC1 (10 ml per step). Fractions (4 ml) were collected and assayed for IEMT activity. Fractions with the highest IEMT activity were pooled (total of 12 ml) and dialyzed against 800 ml ice-cold buffer A overnight. DEAE-cellulose-purified fractions were loaded on a hydroxyapatite column (1-cm-diameter × 10-cm, Bio-gel HT; Bio-Rad, Richmond, CA) installed in a Pharmacia FPLC apparatus and pre-equilibrated with buffer A. Flow rate was 0.3 ml/min. After the enzyme was loaded, the column was washed with 20 ml buffer A and eluted with a linear gradient (100 ml) from 0 to 400 mM Na Phosphate in buffer A. Fractions (2 ml) were collected and assayed for IEMT activity.

The fractions containing IEMT activity, eluted from the HAP column with about 80 mM phosphate, were pooled (7 ml) and loaded immediately onto an adenosine-agarose affinity column (1 ml, Sigma), prepared by modifying the procedures described by Attieh *et al.* (1995). The column was washed with 3 ml buffer A containing 0.3 M NaCl, then with 3 ml buffer A. Finally, IEMT was eluted with 10 ml of 5 mM SAM solution in buffer A. Fractions (1 ml each) were collected and protein content and purity were examined by SDS-PAGE gel electrophoresis followed by silver staining of the gel. The fractions which had the largest amount of pure protein (4 ml) were dialyzed against buffer A using Ultrafree-IS centrifugal filter device (Millipore Corp., Bedford, MA) until the SAM concentration decreased to < 0.5 μM. The purified protein was used for protein N-terminal sequencing, internal peptide mapping and initial enzyme characterization.

Protein Sequencing

The two proteins present in the final purified OMT preparation were separated on long SDS-PAGE and subjected to N-terminal sequencing in a protein sequencer (model 477, Applied Biosystems, Foster City, CA). In addition, the proteins were cleaved with cyanogen bromide, the digestion products were subjected to SDS-PAGE, and additional peptides were isolated and sequenced as previously described (Dudareva *et al.*, 1996).

Isolation and Characterization of cDNA Clones

Since peptide sequence determination showed that the putative IEMT had substantial sequence similarity to COMT, a done encoding *Populus tremuloides* (aspen) COMT (Bugos *et al.*, 1991) was obtained from Dr. W. Campbell (Michigan Technological University, Houghton) and used to screen a *C. breweri* flower cDNA Ubrary (Dudareva *et al.*, 1996). Clones were identified, isolated, and characterized as previously described (Dudareva *et al.*, 1996).

RNA Isolation and RNA Gel-Blot Analysis

Total RNA was isolated from 0.1 g of frozen plant tissue as previously described (Dudareva *et al.*, 1996). RNA samples (7 μg per lane in blots to determine tissue-specific expression, 3 μg per lane in blots to determine variation in expression in petals over the lifespan of the flower) were size-fractionated by electrophoresis under denaturing conditions in vertical urea-agarose gels at 4°C for 5 h at 20 W and transferred to Hybond-N^+ membranes (Amersham, Arlington Heights, IL). For IEMT blots, either a 13-kb IEMT cDNA fragment containing the coding region of the gene or a 0.3-kb fragment containing the 3′ noncoding region of the gene (from nucleotide 1159 in FIG. 6 to the 3′ end) was used as a probe. For COMT blots, a 1-kb fragment containing the coding region of aspen COMT was used as a probe. Hybridizations were performed in 5× SSPE (0.75 M NaCl, 0.05 M NaH_2PO_4, pH 7.4, and 0.5 mM EDTA), 50% formamide, 5× Denhardt's solution, and 0.5% SDS at 37°C for at least 18 h. Membranes were washed at 37°C with 5× SSPE, 0.5% SDS, and twice with 2X × SSPE at 65°C before being exposed to x-ray film. mRNA transcripts were quantified using a molecular imaging system (model GS-363, Bio-Rad, Hercules, CA). IEMT mRNA transcript levels were

normalized to rRNA levels to overcome error in RNA quantitation by spectrophotometry. In addition, all gels contained a standard for equating signals among gels.

Expression of IEMT in Escherichia coli

The cDNA clone of IEMT was subcloned into the *Nde*I-*Bam*HI sites of the pET-T7-11a expression vector (Studier *et al.*, 1990) by first amplifying, by PCR™, the entire coding region of the clone with an oligonucleotide at the 5′ end that introduced an *Nde*I site around the first ATG codon.

A pET-T7 (11a) (Studier *et al.*, 1990) plasmid with the IEMT coding region subcloned into the *Nde*I-*Bam*HI sites (Wang *et al.*, 1997) was transformed into *E. coli* BL21 cells. Cells were grown in a rich LB medium (16 g trypton, 10 g yeast extract and 5 g NaC1 in 1 liter medium) with 1 mg/ml ampicillin and 0.5 mg/ml chloroamphenicol at 37°C. When the culture reached OD_{600} of 0.6, IPTG was added to make a final concentration of 0.2 mM and the culture was incubated at 25°C and shaken at 150 RPM for 20 h. This procedure yields a large amount of IEMT in a soluble form, while only a small fraction of IEMT forms insoluble inclusion bodies. The cells were harvested and sonicated in lysis buffer (10 mM NaC1, 50 mM Tris-HCl, 10 mM EDTA, 10% glycerol and 10 mM β-mercaptoethanol, pH 8.0). After spinning cell lysate at 12,000 g for 10 min to pellet cell debris, soluble IEMT was collected in the supernatant fraction. IEMT was purified by the adenosine-agarose affinity column as described above.

Molecular mass estimation

The native molecular mass of IEMT was determined using a Biological Chromatography System (Bio-Rad, Richmond, CA). Purified IEMT was run on a gel filtration column QC-PAK TSK GFC 300GL (Tosohaas, Montgomeryville, PA) at a flow rate of 2 ml/min in a mobile phase of 50 mM Hepes, pH 7.4, 150 mM NaC1 and 5 mM DTT. Fractions (0.5 ml) were collected for enzyme activity assays. The column was calibrated with the following protein standard: cytochrome C (12.4 kD), ovalbumin (45 kD), bovine serum albumin (67 kD), alcohol dehydrogenase (150 kD), IgG (169 kD), β-amylase (200 kD) and

apoferritin (443 kD). A standard curve was obtained by plotting retention time of the protein standards against the log of the molecular weight. The subunit molecular weight was estimated by denaturing PAGE. The SDS-PAGE was carried out on a 13% polyacrylamide gel and calibrated with molecular weight standards in the range of 14.3 - 200 kD (GibcoBRL, Grand Island, NY). Proteins were visualized by silver staining.

Temperature effect on IEMT stability

IEMT purified from *E. coli* were incubated at temperatures ranging from 4°C to 65°C for 30 min, then chilled on ice. Samples incubated at each temperature were then used for enzyme assays. Two independent assays were performed for each point, then an average was taken.

pH optimum of IEMT activity

The optimum pH for IEMT activity was determined using two buffer systems. Reactions were carried out in 50 mM Bis-Tris buffer with pH ranging from 6.0 to 7.0 and 50 mM Tris-HCl buffer with pH ranging from 7.0 to 10.0. Final results are an average of three independent assays.

Effectors: Enzyme assays were performed with one of the following chemicals presented in the assay buffer at the final concentrate of 25 mM: Ca^{2+}, Cu^{2+}, EDTA, Fe^{2+}, Fe^{3+}, K^{+}, Mg^{2+}, Mn^{2+}, Na^{+}, NH_4^{+} and Zn^{2+}. Except for Cu^{2+}, Fe^{2+} and Fe^{3+}, which precipitate under reducing condition, all assay buffers also contained 10 mM β-mercaptoethanol. Results presented are the average of two studies.

Construction of IEMT/COMT hybrid genes and their expression in E. coli

Regions of the aspen COMT cDNA (Bogus *et al.*, 1991) were amplified by PCR™ and used to substitute the homologous regions from the *C. breweri* IEMT cDNA. Substitutions were done by taking advantage of naturally occurring restriction sites in equivalent positions in both genes, or by engineering such sites without changing the coding information, with the exception of position 218, which is changed into Thr in the chimerical proteins, compared to

Ser in IEMT and Asn in COMT. The chimerical gene constructs were designed in the following way: I/C stands for chimerical protein of IEMT/COMT, and the two numbers following the designation I/C indicate the beginning and end of the amino acid residues of an IEMT region which was substituted by the equivalent region from the aspen COMT cDNA.

Determination of K_mS, K_{cat} and K_{cat}/K_m

In all kinetics studies, appropriate enzyme concentration and incubation times were chosen so that the reaction velocity was linear during the incubation time period. To measure the K_m for each substrate, one substrate concentration was fixed at saturated level and the concentration of the other substrate to be measured varied. Lineweaver-Burk plots were made to obtain the K_m and V_{max} values.

EXAMPLE 2

Temporal Variation in Scent Emission by Intact Flowers

The inventors have previously shown (Raguso and Pichersky, 1995) that eugenol, isoeugenol, methyleugenol, and methylisoeugenol are constituents of the scent of *C. breweri* flowers, although some *C. breweri* plants were found that did not produce methyleugenol and methylisoeugenol (Raguso and Pichersky, 1995). To determine the amount of these compounds emitted at different stages of floral development, the inventors performed time-course headspace collections at 12-h intervals, followed by GC-MS analysis, using an inbred line that emits all four compounds. The inventors began headspace collection with buds on the evening before they opened and ended it 4 days later. Little or no emission was detected from unopened flowers (buds). The amounts emitted from buds and open flowers 1 to 4 days after anthesis are shown in FIG. 1A to FIG. 1D. Emissions of eugenol (FIG. 1A) and methyleugenol (FIG. 1C) both peaked on day 2 and declined thereafter. Emissions of isoeugenol (FIG. 1B) and methylisoeugenol (FIG. 1D) showed similar patterns, although peak emission occurred approximately 12 h later than the eugenol and methyleugenol peaks. Emission levels at peak time ranged from 0.22 μg/flower for isoeugenol to 1.5 μg/flower for eugenol per 12 h.

EXAMPLE 3

Localization and Quantification of Phenylpropanoid Emission from the Different Parts of the Flower

To determine the specific parts of the *C. breweri* flowers that emit these phenylpropanoids, the inventors performed studies in which living flowers were modified by selectively excising floral parts, so that only one class of major floral organ (petals, stamens, or pistil) remained attached to the hypanthium and sepals. The inventors then collected headspace volatiles from these modified flowers over a 24-h period. The data obtained were used to calculate the contribution of each part to the total emission of the flower (FIG. 2). These data revealed that the petals were the organs responsible for the majority of the phenylpropanoid emission. However, the emission of eugenol and isoeugenol in such flowers was greatly decreased (in the case of isoeugenol to below-detection levels). This observation suggests that tissues other than petals may be involved in controlling the flux of the pathway, perhaps by supplying precursors. Alternatively, the injury sustained by the flowers in these studies may have influenced the outcome.

EXAMPLE 4

IEMT and COMT Activities In Flowers

IEMT Activity in Flower Parts and Its Temporal Variation

The biochemical steps that lead to eugenol and isoeugenol synthesis, and their immediate precursors, are currently undetermined. However, it appeared likely that methyleugenol and methylisoeugenol could be synthesized from eugenol and isoeugenol, respectively, by the addition of a methyl group to the *p*-hydroxyl (4′-hydroxyl) moiety ~ (FIG. 3). To the knowledge of the inventors, a specific enzyme that methylates the 4′-hydroxyl of isoeugenol has not been previously described. However, methylation of the 3′ hydroxyl of similar compounds has been found to be catalyzed by enzymes collectively termed OMTs that use SAM as the methyl donor (Attieh *et al.*, 1995; Meng and Campbell, 1996). Therefore, the inventors devised an enzymatic assay to test for the presence of IEMTs. The crude extracts were incubated with either eugenol or isoeugenol and [^{14}C]SAM, and the product was extracted

and analyzed. Products were identified by co-migration with standards on TLC plates and by GC-MS analysis.

C. breweri plants that did not emit methyleugenol and methylisoeugenol did not contain any IEMT activity in their floral parts. *C. breweri* plants that did emit these two compounds contained substantial IEMT in the petals, and some activity was also found in stamens and styles, with trace activity found in stigmata (FIG. 4A and FIG. 4B). None of the remaining floral parts, sepals and ovaries, nor leaves were found to contain any IEMT activity.

Whereas some IEMT activity was already found in mature buds just before anthesis, IEMT activity levels increased quickly after the flower opened and reached 90 to 100% of maximal levels on day 1 of anthesis (FIG. 4A and FIG. 4B). Subsequent IEMT activity remained fairly constant. Interestingly, the patterns of changes in eugenol-OMT activity levels and isoeugenol-OMT activity levels were very similar (FIG. 4A and FIG. 4B), with the ratio of the latter to the former ranging from 1.4 to 1.6 in all floral tissues and throughout the lifespan of the flower.

EXAMPLE 5

COMT Activity in Flower Parts and Its Temporal Variation

The inventors also measured the activity in floral parts of COMT, an enzyme involved in lignin biosynthesis and possibly in isoeugenol biosynthesis as well. Similar to IEMT, COMT activity was highest in petals, but its pattern of gradual increase in activity over time was substantially different from the pattern for IEMT. Whereas IEMT increased quickly after anthesis and reached a maximum on the second day after anthesis, COMT activity was relatively low at this time and increased only gradually to achieve a maximum on the fifth day after anthesis, at the end of flower's lifespan.

EXAMPLE 6

Isolation and Characterization of IEMT cDNA Clones

The inventors purified IEMT from one- to four-day-old petal tissue in a procedure involving anion-exchange, hydroxyapatite, and adenosine-conjugated affinity column chromatography. The latter column is known to selectively bind a variety of OMTs (Attieh *et al.*, 1995). This procedure yielded two proteins that migrated very close to each other on SDS-PAGE, with apparent molecular mass of 40 kD (FIG. 5, right lane). After separating the two proteins on long SDS-PAGE and subjecting them to N-terminal protein sequencing, the inventors obtained two very similar peptide sequences (from the top protein band: XTGNAETQLTP [SEQ. ID NO:3, X = unidentified]; from the bottom protein band: SPGNAEIQIIP [SEQ. ID NO:4]), both of which show some similarity to dicotyledonous plants COMT N-terminal sequences in the data bank. In addition, several individual peptides were obtained from SDS-PAGE after cyanogen bromide cleavage of a mixture of the two proteins, and their sequences also showed significant similarity to COMT sequences. None of the peptide sequences showed any similarity to CCOMT, an enzyme that methylates caffeic acid bound to CoA in an alternative lignin biosynthesis pathway and that also has no significant similarity to COMT (Ye *et al.*, 1994).

Since the N-terminal sequences of both proteins in the "purified" IEMT preparation showed similarity to COMT, the inventors used a COMT cDNA clone from aspen (Bugos *et al.*, 1991) as a probe in low-stringency hybridization screening of a *C. breweri* flower cDNA library. Several clones were isolated, and the nucleotide sequence of one of them, designated IEMT1, was determined for this study (FIG. 6). This clone contains 1486 nucleotides, not including the poly(A) tail, with an open reading frame of 368 codons, beginning with an ATG codon at positions 43 to 45. There is one stop codon in-frame upstream of this ATG, suggesting that the isolated cDNA clone contains the entire coding region. This conclusion is also supported by primer-extension studies to determine the 5′ end of the mRNA. The molecular mass of the protein encoded by the open reading frame of IEMT1 is 40 kD, the same as that of the two proteins found in the purified OMT preparation (FIG. 5).

The N-terminal sequence of the protein encoded by IEMT1 does not match the N-terminal sequence of the higher-molecular-mass protein of the two present in the purified OMT preparation (3/10 mismatches). However, it did match in 9 of 10 positions the N-terminal sequence of the lower-molecular-mass protein (underlined sequences in FIG. 6). The one mismatch in the sequence may have been due to the difficulty in determining the correct N-terminal peptide sequence due to high background signals. The N-terminal sequence predicted from the open reading frame of IEMT1 is two amino add residues longer than that of the N terminus experimentally determined, suggesting that some processing occurs either *in vivo* or during the purification procedure. In addition, the protein encoded by IEMT1 contains two of the internal peptide sequences determined experimentally, MLDRVLRLLASYSVVTYTLRE (SEQ. ID NO:5) and MFDGVPKGDAIFIK (SEQ. ID NO:6) (FIG. 6). The IEMT1 protein is approximately 30% divergent from all available dicot COMT-sequences. These COMTs vary among themselves by no more than 15%, with the exception of a *Zinnia elegans* sequence, designated COMT, that has a substrate specificity that has not been extensively tested (Ye and Varner, 1995). IEMT1 shows sequence identity of 65% or less to several other types of plant OMTs in the data bank, and it contains the three conserved motifs (FIG. 6) identified by Kagan and Clarke (1994) and hypothesized by these authors to be involved in the binding of SAM.

To determine the enzymatic activity of the protein encoded by IEMT1, the inventors cloned it into a pET-T7 (11a) expression system (Studier *et al.*, 1990) and expressed it in *E. coli*. Cell lysates were prepared and assayed with the substrates eugenol and isoeugenol as well as with several intermediates in the lignin biosynthesis pathway that are structurally related to isoeugenol, including caffeic acid. The inventors also tested the substrate specificity of cloned plant COMT expressed in *E. coli* (Meng and Campbell, 1996) and the substrate specificity of the *C. breweri* affinity-purified OMT preparation (FIG. 5, right lane). The latter preparation showed high levels of OMT activity with eugenol, isoeugenol, caffeic acid, and 5-hydroxyferulic acid (FIG. 7A). However, lysates of *E. coli* expressing COMT had high activity only with caffeic add and 5-hydroxyferulic add, as previously described (Meng and

Campbell, 1996), and essentially no activity with eugenol and isoeugenol (FIG. 7C). Lysates of *E. coli* expressing IEMT1 had high levels of activity with both eugenol and isoeugenol, but little or no activity with any of the other substrates tested (FIG. 7B). Moreover, the ratio of isoeugenol-OMT activity to eugenol-OMT activity in the IEMT-expressing *E. coli* lysates (1.4) was basically the same as that of the plant-purified OMT preparation and of the plant crude extracts as well (FIG. 4A and FIG. 4B). Lysates of *E. coli* that contained a pET-T7 (11a) plasmid with no plant DNA insert had undetectable OMT activity with any of these substrates.

Since the inventors' results indicate that IEMT has only residual COMT activity (<3%) and that COMT has only residual IEMT activity, it appears that the substrate-specificity profile of the affinity-purified plant preparation (FIG. 7A) was obtained because this preparation has both COMT and IEMT activities each contained on a separate protein. The lack of activity of COMT with eugenol and isoeugenol further indicates that the IEMT activity measurements shown in FIG. 4C indeed represent only the activity of a distinct enzyme, *i.e.*, IEMT. This enzyme was identified from the N-terminal sequence data as the lower-molecular mass protein in the purified OMT preparation (FIG. 5).

EXAMPLE 7

Northern Blot Analysis of IEMT in Flower Parts

Levels of IEMT mRNA in the different tissues of methylisoeugenol emitters and nonemitters were determined by Northern blots (FIG. 8A and FIG. 8B). mRNAs were extracted from the different tissues of flowers on the day of anthesis. In methylisoeugenol emitters, the highest levels of IEMT mRNA were observed in petals, followed by style and stamens. Little or no IEMT mRNA was observed in stigma and sepals, as well as in leaf and stem tissue (FIG. 8A). It should be pointed out that although IEMT mRNA levels are higher in style than in stamens on a per cell basis, the style is a much smaller organ than stamens are, and as a consequence there is much less total IEMT mRNA in style than in stamens, consistent with the activity data shown in FIG. 4A and FIG. 4B. True-breeding plants that do not emit methylisoeugenol had little or no IEMT message in any of the tissues examined (FIG. 8B).

The size of the IEMT mRNA was estimated to be 1.6 kb, in agreement with the size of the cDNA clone. A broad band, and in many lanes two closely spaced bands, were observed, indicating some heterogeneity in message size. A probe obtained from the 3′ noncoding region of IEMT1 also gave the same pattern. The appearance of two distinct bands is most likely due to the fact that IEMT message fortuitously co-migrates with the abundant 185 rRNA, so that IEMT mRNA transcripts that are slightly larger than 185 rRNA, and IEMT mRNA transcripts that are slightly smaller than 185 rRNA, are separated by the 185 rRNA. This interpretation is also consistent with Southern blot analysis suggesting that methylisoeugenol emitting *C. breweri* plants (as well as non-emitting plants) have a single IEMT gene. With a probe derived from the coding region of COMT, a single band of 13 kb was observed, indicating that under the inventors' stringent Northern blot conditions the IEMT probe does not cross-hybridize with COMT message but only with its own distinct message.

The variation in IEMT mRNA levels over the lifespan of the methylisoeugenol-emitting flowers was also examined. mRNA levels increased as the bud matured and peaked just before anthesis. IEMT mRNA levels on the day of anthesis were about 75% of the levels just before anthesis, and remained relatively stable afterward (FIG. 9A and FIG. 9B).

The inventors shown that the initial synthesis of phenylpropanoid scent compounds in *C. breweri* occur in the same part of the flower from which such compounds are emitted. Although evidence of enzymatic activity in flower tissue responsible for new synthesis of scent compounds is still scarce, this report contributes additional evidence that, scent compounds are produced *de novo* in the tissues from which they are emitted. The inventors conclude that the difficulties encountered in the past in identifying scent-volatile-producing enzymes will be overcome as assays are developed to identify these enzymes, and are probably not due to the absence of such enzymes from floral tissues.

EXAMPLE 8

Biochemical Characterization of (iso)eugenol-o-methyltransferase

IEMT purification

The examples above show the partial purification of IEMT from crude extract of *C. breweri* petals and the isolation of a cDNA clone encoding IEMT (see also Wang *et al.*, 1997). The following is a brief summary of the purification scheme. After chromatography through a DEAE-cellulose anion exchange column, an hydroxyapatite column, and a dephosphorylated adenosine-agarose column, a preparation that contained two proteins of approximately 40 kD in mass was obtained (Table 2 and FIG. 11). With a starting material of 125 ml of crude extract of petals, a typical preparation yielded 16 μg of the larger molecular mass protein and 24 μg of the smaller molecular mass protein. The mixture of the two proteins had O-methyltransferase activity with eugenol, isoeugenol, caffeic acid and 5-hydroxyferulic acid, as previously reported (Wang *et al.*, 1997).

TABLE 2

Purification of IEMT from *Clarkia breweri*

Purification Step	Total Protein (mg)	Total Activity (pKat)	Specific Activity (pKat/mg)	Recovery (%)	Purification (-fold)
Crude extract[a]	98.20	278.8	2.8	100.0	1.0
DE-53	5.51	93.4	17.0	33.5	6.0
Hydroxyapatite	2.13	77.0	36.2	27.5	12.7
Adenosine-agarose	0.04	44.3	1107.5	16.2	397.1

[a]The starting material was 12.5g of petal tissue.

For the studies reported here, the inventors expressed the *C. breweri* IEMT cDNA in *E. coli* to produce a soluble protein. After a single-step purification procedure using the adenosine-agarose affinity column, a pure IEMT protein preparation was obtained (FIG. 11). This protein has O-methyltransferase activity only with eugenol and isoeugenol as substrates.

It comigrates on SDS-PAGE with the lower molecular mass protein in the preparation obtained from plant crude extract (FIG. 11), confirming that this protein is IEMT, as previously inferred from comparing the N-terminal sequence of this protein with the deduced N-terminal sequence of the IEMT cDNA clone (Wang *et al.*, 1997). The upper band is COMT, whose sequence is substantially similar to that of COMT from other plant species (Wang and Pichersky, 1997).

The specific activity of plant-purified IEMT was determined to be 1107.5 pkat/mg protein (pkat = pmol product formed/sec). Since the plant-purified IEMT is a mixture of IEMT and COMT, and IEMT constitutes about 60% of the total protein purified (FIG. 11), the correct specific activity of plant-purified IEMT can be calculated as 1845.8 pkat/mg. The specific activity of IEMT produced in, and purified from, *E. coli* was found to be 2083.3 pkat/mg. Thus, the plant-purified IEMT had a slightly lower specific activity than that of *E. coli-purified* IEMT. The loss of 11% in specific activity in the plant-purified IEMT may be due to the multiple steps in the protein purification procedure, whereas IEMT from *E. coli* was purified by a one-step procedure. Thus, further characterization of plant IEMT was performed with the enzyme purified from *E. coli.*

Tertiary structure of IEMT

The native molecular mass of active IEMT was determined by gel filtration chromatography to be 88 kD (FIG. 12). Since the molecular mass of IEMT determined on SDS-PAGE gel and also calculated from the complete protein sequence is 40 kD, the inventors conclude that the native IEMT is a homodimer.

Temperature effect on IEMT stability

IEMT was incubated at a broad range of temperature for 30 min followed by activity assay. The results show that IEMT was stable at temperature up to 35°C. After 30 min incubation at 45°C, its activity was dramatically decreased (FIG. 13).

pH optimum

The pH dependence of IEMT activity was examined in the pH range of 6.0-10.0, using buffers of Bis-Tris and Tris with overlapping ranges. IEMT exhibited a pH optimum at 7.5 for both eugenol and isoeugenol methylation (FIG. 14).

Co-factor effect

The effect of various co-factors on the enzymatic activity of IEMT were tested (Table 3). No co-factor requirement for IEMT was detected; activity without any co-factor was as high or higher than with any co-factor tested. Positively charged ions such as K^+, Na^+, NH_4^+ and Ca^{2+} showed only weak inhibitory effect on IEMT activity. Mg^{2+} had a moderate inhibitory effect on IEMT activity. The heavy metals Cu^{2+}, Fe^{2+}, Fe^{3+}, Zn^{2+} had strong inhibitory effect on IEMT activity.

K_m for SAM, eugenol* and *isoeugenol

The reactions catalyzed by IEMT exhibited Michaelis-Menten kinetics with respect to its substrate saturation responses. The apparent K_m values for the isoeugenol, eugenol and SAM were determined to be 74 μM, 30 μM and 19 μM, respectively.

TABLE 3

Co-factor Effect on IEMT Activity

Effector[a]	Relative Activity (%)
None	100
Na+	90±4[b]
Ca^{2+}	85±2
NH_4^+	85±12
K^+	83±3
Mg^{2+}	67±8
Mn^{2+}	23±8
Fe^{2+}	18±6
Fe^{3+}	5±1
Cu^{2+}	3±1
Zn^{2+}	3±1

[a]All effector molecules were tested at a concentration of 5 mM.

[b]Results reported are the average of two assays.

From these characterization studies it can be surmised that IEMT shares many properties with COMTs from aspen and other plant species. Similarly to COMT, IEMT is a homodimer with a subunit molecular weight of 40 kD. Its optimal pH of 7.5 and its temperature stability profile are also similar to that of COMTs. Also similarly to COMTs, IEMT exhibits no requirement for divalent cations for its activity, but is strongly inhibited by heavy metals such as Cu^{2+}, Fe^{2+}, Fe^{3+}, Mn^{2+} and Zn^{2+}. Among plant SAM-dependent OMTs in general, some are enhanced by Mg^{2+} (Benavente-Garcia *et al.*, 1995), while others are not affected (Wanek and Richter, 1995; Sato *et al.*, 1994; Hara *et al.*, 1995).

The K_ms values of 74, 30 and 19 μM for isoeugenol, eugenol and SAM, respectively, were comparable to the K_ms of 40 μM for caffeic acid and 6.2-8.4 μM for SAM determined for aspen COMT (Meng and Campbell, 1996). K_m values for SAM of other plant

O-methyltransferases were found to be between 10-100 μM (Benavente-Garcia *et al.*, 1995; Wanek and Richter, 1995; Maxwell *et al.*, 1992; De Carolis and Ibrahim, 1989).

EXAMPLE 9

Localization of the regions in IEMT involved in substrate binding

Three sequence "motifs" which are relatively highly conserved in all methyltransferases that use SAM as the methyl group donor have been identified (Schluckebier *et al.*, 1995; Kagan and Clarke, 1994). Moreover, the three-dimensional structure of catechol O-methyltransferase shows that these regions interact with SAM (Vidgren *et al.*, 1994). In plant OMTs, these motifs are all found in the C-terminal half of the protein (Kagan and Clarke, 1994). Although IEMT and COMT sequences are 73-83% identical (FIG. 15) and have similar catalytic activities (the methylation of an hydroxy group), they bind different substrates. To test for the specific regions in IEMT that participate in binding eugenol/isoeugenol and in catalysis, the inventors constructed several IEMT/COMT chimerical genes in which a specific region in the N-terminal half of IEMT was replaced with the equivalent coding region from aspen COMT (FIG. 16). These constructs were expressed in *E. coli* and resulting enzymes were tested for activities with caffeic acid, 5-hydroxyferulic acid, eugenol and isoeugenol. These enzymes were further purified from *E. coli,* and the kinetic parameters K_m, (binding affinity), K_{cat} (specific activity), and K_{cat}/K_m were determined. The latter parameter is an indication of the catalytic efficiency (Voet and Voet, 1995).

The results are summarized in FIG. 16. Substituting as little as 116 amino acids from COMT (I/C92-207) resulted in an enzyme that methylated caffeic acid and 5-hydroxyferulic acid but not eugenol or isoeugenol, with a K_m for caffeic acid similar to that of the original aspen COMT. The catalytic efficiency of I/C92-207 was 2/3 of that of the original COMT. Substituting a smaller fragment of 82 amino acids (I/C92-173) also resulted in an enzyme which was active with caffeic acid and 5-hydroxyferulic acid, but not with eugenol or isoeugenol. However, the K_m for caffeic acid of I/C92-173, 155 μM, was 5 times higher than the K_m of I/C92-207, indicating that I/C92-173 has 5 times weaker affinity to caffeic acid than does I/C92-207. Thus, although the specific activity of I/C92-173 was 2/3 that of the original

COMT, its overall catalytic efficiency was only 18% that of original COMT and 27% of I/C92-207.

Substituting other parts of the N-terminal region of IEMT with equivalent COMT sequences (I/C46-91 and I/C173-207) did not change the substrate specificity of the protein, which still had activity with eugenol and isoeugenol, but not with caffeic acid nor S-hydroxyferulic acid (FIG. 16). Moreover, the resulting enzymes had K_ms for isoeugenol (65 μM and 64 μM, respectively,) similar to the K_m of the original IEMT (74 μM), indicating that these flanking regions did not have significant effect on IEMT substrate binding. However, I/C46-91 had a much reduced specific activity (25% of authentic IEMT), and consequently its catalytic efficiency was also much lower than the original IEMT. The specific activity and catalytic efficiency of I/C173-207 were similar to the original IEMT.

Discussion

When most of the N-terminal half of IEMT was substituted by the equivalent fragment from COMT, the chimerical protein (FIG. 16, I/C92-207) had catalytic activity with caffeic acid, but not with (iso)eugenol. The K_m value for caffeic acid of this protein was similar to the K_m value of authentic COMT. This result suggested that the substrate specificity domain of IEMT and COMT resides in the first half of protein between amino acid 92-207. Further "domain replacement" studies showed that an IEMT protein with only 82 amino acid fragment replaced by the corresponding region from COMT (I/C92-173) still was specific for caffeic acid but not (iso)eugenol, although its catalytic efficiency was only 18% of the original COMT. The overall loss of its catalytic efficiency was attributed mostly to its loss of affinity for substrate binding (it has a 5-fold larger K_m value). Thus, The flanking region between position 173-207 appears to play a role in modulating the binding of caffeic acid.

When fragments flanking the 92-173 region of IEMT were replaced by the corresponding fragments from COMT (FIG. 15C, I/C46-91 and I/C173-207), both chimerical proteins showed IEMT activity. This is consistent with the inventors' previous results indicating that segment 92-173 is the main region in IEMT and COMT to determine both

substrate specificity and the position of the hydroxy group to be methylated. An IEMT protein with both flanking regions replaced with COMT sequences (I/C46-91,173-207) also acted as IEMT. While hybrid protein I/C173-207 was essentially identical in its kinetic properties to authentic IEMT, I/C46-91 had a K_m for isoeugenol similar to that of IEMT, but its specific activity was only 25% of that of IEMT, resulting in overall 72% loss of its catalytic efficiency. These results suggested that: 1) region 173-207 may not be important for (iso)eugenol binding and IEMT catalytic activity (although it does have an effect on caffeic acid binding, see above), and 2) region 46-91 of IEMT also does not affect substrate binding but does impair IEMT activity. The inventors conclude that the segment which accounts for the main differences between IEMT and COMT in both substrate specificity and the position to be methylated resides in the segment extending from position 92 to 173, a region of 82 amino acids.

It is interesting that the inventors were not able to separate the substrate specificity function from the specificity of methylation position, and that both specificities seem to reside in region 92-173. Moreover, since in this region there are only 16 positions where the amino acid in IEMT is a different one from that always found in COMT sequences (FIG. 15), the differences in substrate specificity and methylation position between COMT and IEMT could be accounted for by no more than these 16 amino acids, or fewer. However, as discussed above, the adjacent regions also contribute to the substrate binding and catalytic efficiency. These observations could perhaps be explained if region 92-173 forms a core substrate binding pocket, which binds the substrate and present it to the active site, while its flanking regions 46-91 and 173-207 are important in stabilizing the protein-substrate complex and also in affecting the catalytic process.

(i) IEMT evolution

The overall sequence of IEMT from *Clarkia breweri* shows 70-83% sequence identity with COMT, indicating that they arose from a common ancestral gene. Although *C. breweri* IEMT has the highest sequence identity to *C. breweri* COMT (83%) than to any other COMT sequence in the databank, *C. breweri* COMT still has higher sequence similarity (84%) to some COMT sequences from related species within the class Rosidae (Wang and Pichersky, 1997).

This observation suggests that the gene duplication that gave rise to IEMT occurred soon after the creation of the Rosidae lineage. Alternatively, IEMT evolved from COMT only recently but has diverged at a higher rate than COMTs are diverging from each other.

Methyleugenol and isomethyleugenol, as well as estragol and methylchavicol (analogs of methyleugenol and isomethyleugenol which lack a side chain at the 3′ position), are all synthesized in plants by the methylation of the 4′ OH group of the benzyl ring (Wang *et al.*, 1997; De Carolis and Ibrahim, 1989). Various plant species produce one or more of these compounds. For example, methylchavicol constitutes the bulk of the essential oil fraction of some basil varieties (Simon *et al.*, 1990). Unfortunately, data on the enzymes that catalyze the formation of these four methylated compounds in different plant species is lacking. However, the inventors have found that IEMT from *C. breweri* cannot methylate chavicol. Since it appears that new substrate specificities and positions of methylation can evolve by relatively few changes from an existing OMT, such enzymes could have evolved independently of each other. Thus, whether all 4′ OH methyltransferases from plants have evolved once from the COMT lineage or more than once remains an open question, which could be resolved with the availability of additional sequences of such enzymes.

EXAMPLE 10

In Vitro Mutagenesis of IEMT

The genes for a single amino acid mutant at a given position may be constructed by overlap PCR, then corresponding mutant proteins are synthesized in coupled *in vitro* transcription-translation reactions. For each mutant, the relative enzyme activity may be determined. Techniques for performing such mutations are well know to those of skill in the art (*e.g.* Burks *et al.*, 1997). Meng and Campbell (1996) describe site-directed mutagenesis of a lignin-specific O-methyltransferase, inasmuch as the methodology described therein could be employed to produce mutants of IEMT the entire text is incorporated by reference. In order to form such mutants, primers targeted to mutate particular residues are designed so that the primers encode mutants at given sites.

Using such methods, certain amino acid substitutions were identified that are necessary for optimal IEMT activity (FIG. 17). For example, while the wild-type amino acids between residues 133-135 (TAT) have an IEMT/COMT activity ratio of 33.40, mutating these residues to MNQ respectively reduces the activity ratio to 2.98. Similarly the presence of NE at residues 164-165 is important for IEMT activity as mutating these residues to TA as found in COMT decreases the IEMT/COMT activity ratio to 3.18. A double mutation in which the residues at 133-135 comprise MNQ and the residues at 164-165 comprise TA produces a further decrease in the IEMT/COMT activity ratio to 0.13. A mutation in which the H at residue 173 is mutated to a P decreases the activity to 10.54. Mutating the S102⌫D102 results in a similar decrease in IEMT/COMT activity ratio. Also important for IEMT activity are the residues at 130-131, mutation of FL130-131⌫LC130-131 results in an IEMT/COMT ratio of 9.38. A double mutation containing LC at residues 130-131 and TA at residues 164-165 produced a further decrease in IEMT/COMT activity ratio (to 0.98). Even more of a decrease in IEMT activity was observed in the double mutant containing LC at residues 130-131 and MNQ at residues 133-135 (IEMT/COMT activity ratio 0.05). A triple mutant, in which the wild-type sequence (SEQ ID NO:10) is mutated so that the residues at 130-131, 133-135 and 164-165 are LC, MNQ, TA, produced the most dramatic decrease in IEMT activity implicating the importance of FL, TAT, and NE amino residues at these respective positions in the wild-type IEMT protein.

* * *

All of the compositions and/or methods disclosed and claimed herein can be made and executed without undue experimentation in light of the present disclosure. While the compositions and methods of this invention have been described in terms of preferred embodiments, it will be apparent to those of skill in the art that variations may be applied to the compositions and/or methods and in the steps or in the sequence of steps of the method described herein without departing from the concept, spirit and scope of the invention. More specifically, it will be apparent that certain agents which are both chemically and physiologically related may be substituted for the agents described herein while the same or similar results would be achieved. All such similar substitutes and modifications apparent to

those skilled in the art are deemed to be within the spirit, scope and concept of the invention as defined by the appended claims.

REFERENCES

The following references, to the extent that they provide exemplary procedural or other details supplementary to those set forth herein, are specifically incorporated herein by reference.

Abdullah *et al.*, *Biotechnology*, 4:1087, 1986.

Ackermann *et al.*, *J. Plant Physiol.*, 134:567-572, 1989.

Attieh *et al.*, *J. Biol. Chem.*, 270:9250-9257, 1995.

Barany and Merrifield, The Peptides, Gross and Meienhofer, eds., Academic Press, New York, pp. 1-284, 1979.

Bellus, *J. Macromol. Sci. Pure Appl. Chem*, A31(1): 1355-1376, 1994.

Benavente-Garcia *et al.*, *Plant Physiol. Biochem.*, 33:263-271, 1995

Benbrook *et al.*, *In: Proceedings Bio Expo 1986*, Butterworth, Stoneham, MA, pp. 27-54, 1986.

Bevan *et al.*, *Nucleic Acid Research.*, 11:369-385, 1983.

Bogus *et al.*, *Plant Mol. Biol.*, 17:1203-1215, 1991

Bouchez, *EMBO Journal*, 8(13):4197-4204, 1989.

Braisted *et al.*, *Proc. Natl. Acad. Sci. USA*, 93:5688-5692, 1996

Brown *et al.*, *J. Immunology*, 156:3285-3291, 1996

Bugos *et al.*, *Plant Mol. Biol.*, 17:1203-1215, 1991.

Burks *et al.*, *Proc. Natl. Acad. Sci. USA*, 94:412-417, 1997 .

Burton and Barbas, Adv. Immunol., 3rd, 57:191-280, 1994

Bytebier *et al.*, *Proc. Natl. Acad. Sci. USA*, 84:5345, 1987.

Callis *et al.*, *Genes and Develop.*, 1:1183-1200, 1987.

Campbell, In: *Monoclonal Antibody Technology, Laboratory Techniques in Biochemistry and Molecular Biology*, Burden and Von Knippenberg (Eds.), Elseview, Amsterdam, 13:75-83, 1984.

Capaldi *et al.*, *Biochem. Biophys. Res. Comm.*, 76:425, 1977

Cashmore *et al.*, *Gen. Eng. of Plants*, Plenum Press, New York, 29-38, 1983.

Chandler *et al.*, *The Plant Cell*, 1:1175-1183, 1989.

Chau *et al.*, *Science*, 244:174-181, 1989.

Christou *et al.*, *Plant Physiol*, 87:671-674, 1988.

Chu *et al.*, *Scientia Sinica*, 18:659-668, 1975.

Conkling *et al.*, *Plant Physiol.*, 93:1203-1211, 1990.

Cooley *et al. Science*; 239(4844): 1121-1128, 1988

Cunningham *et al.*, *Science*, 243(4896):1330-1336, 1989.

De Block *et al.*, *EMBO J.*, 6:2513-2518; see also PCT Publication number WO 87/05629, published 9/24/87, 1987.

De Block *et al.*, *Plant Physiol*, 91:694-701, 1989.

De Carolis and Ibrahim, *Biochem. Cell Biol.*, 67:763-769, 1989.

Dellaporta *et al.*, *in Chromosome Structure and Function: Impact of New Concepts, 18th Stadler Genetics Symposium*, J.P. Gustafson and R. Appels, eds (New York: Plenum Press), pp. 263-282, 1988.

Dellaporta *et al.*, *Stadler Symposium*, 11:263-282, 1988.

Dhir *et al.*, *Plant Cell Reports*, 10:97, 1991

Dobson, *In: Insect-Plant Interactions*, Vol. V, E. Bernays, ed., CRC Press, Boca Raton, FL, pp. 47-81, 1993.

Dodson *et al.*, *Science*, 164:1243-1249, 1969.

Dudareva *et al.*, *Plant Cell*, 8:1137-1148, 1996.

Ebert *et al.*, *PNAS*, 84:5745-5749, 1987.

Ellis *et al.*, *EMBO Journal*, 6(11):3203-3208, 1987.

European Patent No. 0189707.

European Patent No. 154 204, 1985

European Patent No. 320 308,

European Patent No. 329 822

Fagan and Clarke, *Arch. Biochem. Biophys.*, 310:417-427, 1994.

Finkle *et al.*, *Plant Sci*, 42:133-140, 1985.

Fodor *et al.*, *Science*, 251:767-773, 1991.

Fraley *et al.*, *Biotechnology*, 3:629, 1985.

Fraley *et al.*, *Proc. Natl. Acad. Sci. USA*, 80:4803, 1983.

Fransz *et al.*, *Plant Cell Rep*, 8:67-70, 1989.

Freifelder, *Physical Biochemistry Applications to Biochemistry and Molecular Biology*, 2nd ed. Wm. Freeman and Co., New York, NY, 1982.

Frohman, M.A., In: *PCR™ protocols: A guide to methods and applications*, Academic Press, N.Y., 1990

Fromm *et al.*, *Bio Technology 8*, 833,839, 1990.

Fromm *et al.*, *Nature*, 312:791-793, 1986.

Fromm *et al.*, *The Plant Cell*, 1:977-984, 1989.

Fujimura *et al.*, *Plant Tissue Culture Letters*, 2:74, 1985.

Galen and Kevan, *Can. J. Zool.*, 61:1207-1213, 1985.

Galen, *Ecology*, 66:792-797, 1985.

Gallie *et al.*, *The Plant Cell*, 1:301-311, 1989.

Gefter *et al.*, *Somatic Cell Genet.*, 3:231-236, 1977.

Gelvin *et al.*, *Plant Molecular Biology Manual*, 1990.

Gingeras *et al.*, PCT Application WO 88/10315,

Goding, 1986, In: *Monoclonal Antibodies: Principles and Practice*, 2d ed., Academic Press, Orlando, Fla., pp. 60-61, and 71-74, 1986.

Hacia *et al.*, *Nature Genetics*, 14:441-447, 1996.

Hara *et al.*, *Phytochem.*, 38:1131-1135, 1995.

Harlow and Lane, "Antibodies: A Laboratory Manual," Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1988

Hess, *Intern Rev. Cytol.*, 107:367, 1987.

Hilton *et al.*, *J. Biol. Chem.*, 271:4699-4708, 1996.

Hinchee *et al.*, *Bio/technol*, 6:915-922, 1988.

Horsch *et al.*, *Science*, 227:1224-1231, 1985.

Hudspeth and Grula., *Plant Mol. Biol.*, 12:579-589, 1989.

Ikuta *et al.*, *Bio. Technol*, 8:241-242, 1990.

Innis *et al.*, *PCR™ Protocols*, Academic Press, Inc., San Diego CA, 1990.

Jackson *et al.*, *J. Immunol.*, 154:3310-3319, 1995

Johnson *et al.*, "Peptide Turn Mimetic" in *BIOTECHNOLOGY AND PHARMACY*, Pezzuto *et al.*, Eds., Chapman and Hall, New York, 1993

Jorgensen *et al.*, *Mol. Gen. Genet.*, 207:471, 1987.

Joulain, *Flavor Fragrance J.*, 2:149-155, 1987.

Kagan and Clarke, *Arch Biochem Biophys* 310(2):417-27, 1994.

Katz *et al.*, *J. Gen. Microbiol.*, 129(Pt 9):2703-2714, 1983.

Keller *et al.*, *EMBO J.*, 8(5):1309-1314, 1989.

Klee *et al.*, *Biotechnology*, 3:637-642, 1985.

Klein *et al.*, *Nature*, 327:70-73, 1987.

Kohler and Milstein, *Eur. J. Immunol.*, 6:511-519, 1976.

Kohler and Milstein, *Nature*, 256:495-497, 1975.

Koncz *et al.*, *EMBO J.* 9(5):1337-1346, 1990.

Kwoh *et al.*, *Proc. Nat. Acad. Sci. USA*, 86: 1173, 1989.

Kyte and Doolittle, *J. Mol. Biol.*, 157(1):105-132, 1982.

Langridge *et al.*, *Proc. Natl. Acad. Sci. USA*, 86:3219-3223, 1989.

Laufs *et al.*, *Proc. Natl. Acad. Sci.*, 7752-7756, 1990.

Lawton *et al.*, *Plant Mol. Biol.*, 9:315-324, 1987.

Lindstrom *et al.*, *Developmental Genetics*, 11:160, 1990.

Lorz *et al.*, *Mol Gen Genet*, 199:178-182, 1985.

Loughrin *et al.*, *Phytochemistry* 31:1537-1540, 1992.

Luo *et al.*, *Plant Mol. Biol. Reporter*, 6:165, 1988.

MacSwain *et al.*, *Univ. Cal. Publ. Entomol.*, 70:1--80, 1973.

Maddock *et al.*, *Third International Congress of Plant Molecular Biology*, Abstract 372, 1991.

Manitto *et al.*, *L. Tetrahedron Lett.*, 17:1567-1568, 1974.

Marcotte *et al.*, *Nature*, 335:454, 1988.

Marks *et al.*, *Symp. Soc. Exp. Biol.* 45:77-87, 1991.

Maxwell *et al.*, *Arch. Biochem. Biophys.*, 293:158-166, 1992.

McCabe *et al.*, *Bio. Tech.*, 6:923-926, 1988.

Meng and Campbell, *Arch. Biochem. Biophys.*, 330:329-341, 1996.

Merrifield, *Science*, 232:341-347, 1986

Miller *et al.*, PCT Application WO 89/06700.

Murakami *et al.*, *Mol. Gen. Genet.*, 205:42-50, 1986.

Murashige and Skoog, *Physiol Plant*, 15:473-497, 1962.

Neuhaus *et al.*, *Theor. Appl. Genet.*, 75:30, 1987.

Odell *et al.*, Nature, 313:810-812, 1985.

Ohara *et al.*, *Proc. Natl Acad. Sci. USA*, 86:5673-5677, 1989.

Oppenheimer *et al. Cell* 67(3):483-493, 1991.

Ow *et al.*, *Science*, 234:856-859, 1986.

Pease *et al.*, *Proc. Natl. Acad. Sci. USA*, 91:5022-5026, 1994.

Pellmyr, *Oecologia*, 78:304-307, 1986.

Pena *et al.*, *Nature*, 325:274, 1987.

Pichersky *et al.*, *Arch. Biochem Biophys.*, 316:803-807, 1995.

Pichersky *et al.*, *Plant Physiol.*, 106:1533-1540, 1994.

Pignon *et al.*, *Hum. Mutat.*, 3: 126-132, 1994.

Poszkowski *et al.*, *EMBO J.*, 3:2719, 1989.

Potrykus *et al.*, *Mol Gen Genet*, 199:183-188, 1985.

Potrykus, *Trends Biotechnol*, 7:269-273, 1989.

Poulsen *et al.*, *Mol. Gen. Genet.*, 205:193-200, 1986.

Prasher *et al.*, *Biochem Biophys Res Commun.* 126(3): 1259-1268, 1985

Raguso and Pichersky, *Plant Syst. Evol.*, 194:55-67, 1995.

Rhodes *et al.*, *Science* 240(4849):204-207, 1988.

Rogers *et al.*, *In: Methods For Plant Molecular Biology*, A. Weissbach and H. Weissbach, eds., Academic Press Inc., San Diego, CA 1988.

Rogers *et al.*, *Meth. in Enzymol.*, 153:253-277, 1987.

Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2d ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

Sato *et al.*, *Eur. J. Biochem.*, 225:125-131, 1994.

Schluckebier *et al.*, *J. Mol. Biol.*, 247:16-20, 1995.

Schmidt *et al. Science* 238(4829):960-963, 1987.

Senanayake *et al.*, *Phytochemistry*, 16:2032-2033, 1977.

Shoemaker *et al.*, *Nature Genetics* 14:450-456, 1996.

Short *et al.*, *J. Biol. Chem.*, 270:28541-28550, 1995.

Simon *et al.*, *In: Advances in New Crops*, Timber Press, Portland, OR, pp 484-489, 1990.

Simpson, *Science*, 233:34, 1986.

Skriver and Mundy, *Plant Cell*, 2(6):503-512, 1990.

Skriver and Mundy, *Plant Cell*, 2:503-512, 1990.

Sommer *et al. EMBO J.*, 9(3):605-613, 1990.

Spielmann *et al.*, *Mol. Gen. Genet.*, 205:34, 1986.

Stalker *et al.*, *Science*, 242: 419-422, 1988.

Stewart and Young, *Solid Phase Peptide Synthesis*, 2d. ed., Pierce Chemical Co., 1984.

Stiefel *et al.*, *The Plant Cell*, 2:785-793, 1990.

Studier *et al.*, *Methods Enzymol.*, 185:60-89, 1990.

Sullivan *et al.*, *Mol. Gen. Genet*, 215:431-440, 1989.

Sutcliffe, *Proc Natl Acad Sci USA*, 75:3737-3741, 1978.

Tam *et al.*, *J. Am. Chem. Soc.*, 105:6442, 1983.

Thill *et al.*, *J. Biol. Chem.*, 263:12500-12508, 1988.

Thompson and Neiman, *Cell*, 48:369-378, 1987.

Toriyama *et al.*, *Theor. Appl. Genet.*, 73:16, 1986.

Twell *et al.*, *Plant Physiol.*, 91:1270-1274, 1989.

U. S. Patent 4,757,011

U. S. Patent 4,769,061

U. S. Patent 4,940,835,

U. S. Patent 4,971,908,

U. S. Patent 5,591,616

U.S. Patent 4,196,265

U.S. Patent 4,554,101

U.S. Patent 4,554,101

U.S. Patent 4,683,195

U.S. Patent 4,683,202

U.S. Patent 4,800,159

U.S. Patent 4,883,750

U.S. Patent 5,279,721

U.S. Patent 5,350,689

U.S. Patent 5,451,514

U.S. Patent 5,489,520

U.S. Patent 5,538,880

Uchimiya *et al.*, *Mol. Gen. Genet.*, 204:204, 1986.

Ugaki *et al.*, *Nucl Acid Res*, 19:371-377, 1991.

Vain *et al.*, *Plant Cell* Reports, 12:84-88, 1993.

Vain *et al.*, *Plant Mol. Biol.*, 18:189-200, 1992.

Van Tunen *et al.*, *EMBO J.*, 7:1257, 1988.

Vasil *et al.*, *Biotechnology*, 10:667-674, 1992.

Vasil *et al.*, *Plant Physiol.*, 91:1575-1579, 1989.

Vasil, *Bio. Tech.*, 8:797, 1990.

Vidgren *et al.*, *Nature*, 368:354-357, 1994.

Vodkin *et al.*, *Cell*, 34:1023, 1983.

Voet and Voet, *In: Biochemistry*, John Wiley and Sons, Inc., p 354, 1995.

Vogel *et al.*, *J. Cell Biochem.*, supplement 13D:312, 1989.

Walker *et al.*, *PNAS*, 84:6624-6628, 1987.

Walker *et al.*, *Proc. Natl. Acad. Sci. USA*, 89:392-396, 1992.

Wanek and Richter, *Planta*, 197:427-434, 1995.

Wang and Pichersky, *Arch Biochem Biophys*. 349(1):153-160, 1998.

Wang *et al.*, *Molecular and Cellular Biology*, 12:3399-3406, 1992.

Wang *et al.*, *Plant Physiology*, 114:213-221, 1997.

Warren *et al.*, *Biochemistry*, 35:8855-8862, 1996.

Watanabe *et al.*, *Biosci. Biotechnol. Biochem.*, 57:1101-1106, 1993.

Weissbach and Weissbach, *Methods for Plant Molecular Biology*, (eds.), Academic Press, Inc., San Diego, CA, 1988.

Wells, *Proc. Natl. Acad. Sci. USA*, 93:1-6, 1996.

Wenzler *et al.*, *Plant Mol. Biol.*, 12:41-50, 1989.

WO 90/07641 filed December 21, 1990.

Wong *et al.*, *J. Bacteriol.*, 178(8):2334-2342, 1996.

Wu *et al.*, *Genomics*, 4:560, 1989.

Xiang. and Guerra, *Plant Physiol.*, 102:287-293, 1993.

Yamada *et al.*, *Plant Cell Rep.*, 4:85, 1986.

Yang *et al.*, *Proc. Natl. Acad. Sci. USA*, 87:9568-9572, 1990.

Yang and Russell, *PNAS*, 87:4144-4148, 1990.

Ye and Varner, *Plant Physiol.*, 108:459-467, 1995.

Ye *et al.*, *Plant Cell*, 6:1427-1439, 1994.

Yelton *et al.*, *J. Immunol.*, 155:1994-2004, 1995.

Zeng *et al.*, *Biochemistry*, 35(40):13157-13164, 1996.

Zhou *et al.*, *Methods in Enzymology*, 101:433, 1983.

Zukowsky *et al.*, *Proc. Natl. Acad. Sci. USA*, 80:1101-1105, 1983.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: The Reagents of the University of Michigan
(B) STREET: 3003 South State Street
(C) CITY: Ann Arbor
(D) STATE: MI
(E) COUNTRY: USA
(F) POSTAL CODE (ZIP): 48109-1280
(G) TELEPHONE: (512) 418-3000

(A) NAME: Eran Pichersky
(B) STREET: 325 Burlington Woods Dr., #180
(C) CITY: Ann Arbor
(D) STATE: MI
(E) COUNTRY: USA
(F) POSTAL CODE (ZIP): 48104

(A) NAME: Jihong Wang
(B) STREET: 1738 Murfin, #21
(C) CITY: Ann Arbor
(D) STATE: MI
(E) COUNTRY: USA
(F) POSTAL CODE (ZIP): 48105

(ii) TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR USE OF (ISO)EUGENOL METHYLTRANSFERASE

(iii) NUMBER OF SEQUENCES: 13

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1486 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

ATAAGTACCA GAAAGCTCTC ATAACAGAAA AAAAAAAAAA AATGGGATC TACCGGAAAT 60

GCAGAGATCC AGATAATCCC CACCCACTCC TCCGACGAGG AAGCCAACCT CTTCGCCATG 120

CAGCTGGCCA GCGCCGCCGT TCTCCCCATG GCCCTTAAGG CCGCCATCGA GCTCGACGTC 180

CTTGAGATCA TGGCCAAGTC CGTCCCTCCC AGCGGCTACA TCTCTCCGGC GGAGATTGCC 240

TCGCAGCTTC CTACCACCAA CCCTGAAGCT CCGGTGATGC TTGACCGTGT CCTCCGCCTC 300

CTAGCCAGCT ACTCCGTCGT AACATACACT CTCCGGGAAC TTCCCAGCGG CAAGGTGGAG 360

AGGCTGTACG GCCTCGCCCC TGTCTGCAAG TTCTTGACCA AGAACGAGGA TGGAGTTTCT 420

CTTGCTCCTT TTTTGCTCAC GGCTACCGAC AAGGTCCTTT TGGAGCCCTG GTTTTACTTG 480

AAAGATGCGA TTCTTGAAGG AGGAATTCCA TTCAATAAAG CGTATGGAAT GAATGAATTC 540

GATTACCATG GAACAGACCA CAGATTCAAC AAGGTGTTCA ACAAGGGAAT GTCCAGCAAC 600

TCTACCATCA CCATGAAGAA GATCCTTGAA ATGTACAACG GATTCGAGGG GCTAACAACG 660

ATTGTCGATG TTGGGGGCGG TACAGGTGCC GTGGCTAGCA TGATTGTTGC TAAGTATCCT 720

TCCATCAACG CCATCAACTT CGACCTGCCT CACGTTATTC AGGATGCTCC AGCTTTTTCT 780

GGTGTTGAAC ATCTTGGAGG AGATATGTTT GATGGCGTAC CCAAAGGCGA CGCTATATTC 840

ATCAAGTGGA TTTGCCACGA CTGGAGCGAT GAGCATTGCC TGAAGTTGCT GAAAAACTGC 900

TATGCTGCAC TTCCCGACCA TGGCAAGGTC ATTGTTGCAG AATACATCCT TCCTCCGTCT 960

CCTGACCCGA GTATCGCCAC CAAGGTAGTC ATCCATACCG ACGCCCTCAT GTTGGCCTAC 1020

AACCCAGGCG GCAAAGAAAG GACTGAGAAG GAGTTCCAGG CTTTGGCTAT GGCTTCCGGA 1080

TTCAGGGGTT TCAAAGTAGC ATCTTGTGCC TTCAACACTT ACGTCATGGA GTTCCTCAAA 1140

ACCGCGTAAA TGATTATGTT CGAAACCGAC CAATTGTGAA TGGCTGCAAA ACTATTCCTA 1200

TCGAATAAGT GAGTTTTATG CTGGTTGTTG CTGAATATAT CAGTATGCAA GAGTATGCTC 1260

TTCCAATAAA TCTTAGAATA GTAGTGACTT TGTACAAGTC CTAGAATAGT GGTAAGCTGT 1320

GTCTTTACTG TTAAAAGTTT GTCGTATGGC CACTATAAAA GGAAAGTATC TGCGTCTTTG 1380

TTGTAATTAG CAATTCACTG TAGCTGAGAT CCTCCCCTCA GCTTAGGTGT TTGCTCTCAA 1440

TTATTCTCCA GCTTAATGTG AATTGAGCCT GACTGGAGCT TATTAG 1486

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 368 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Gly Ser Thr Gly Asn Ala Glu Ile Gln Ile Ile Pro Thr His Ser
1 5 10 15

Ser Asp Glu Glu Ala Asn Leu Phe Ala Met Gln Leu Ala Ser Ala Ala
20 25 30

Val Leu Pro Met Ala Leu Lys Ala Ala Ile Glu Leu Asp Val Leu Glu
35 40 45

Ile Met Ala Lys Ser Val Pro Pro Ser Gly Tyr Ile Ser Pro Ala Glu
50 55 60

Ile Ala Ser Gln Leu Pro Thr Thr Asn Pro Glu Ala Pro Val Met Leu
65 70 75 80

Asp Arg Val Leu Arg Leu Leu Ala Ser Tyr Ser Val Val Thr Tyr Thr
85 90 95

Leu Arg Glu Leu Pro Ser Gly Lys Val Glu Arg Leu Tyr Gly Leu Ala
100 105 110

Pro Val Cys Lys Phe Leu Thr Lys Asn Glu Asp Gly Val Ser Leu Ala
115 120 125

Pro Phe Leu Leu Thr Ala Thr Asp Lys Val Leu Leu Glu Pro Trp Phe
130 135 140

Tyr Leu Lys Asp Ala Ile Leu Glu Gly Gly Ile Pro Phe Asn Lys Ala
145 150 155 160

Tyr Gly Met Asn Glu Phe Asp Tyr His Gly Thr Asp His Arg Phe Asn
165 170 175

Lys Val Phe Asn Lys Gly Met Ser Ser Asn Ser Thr Ile Thr Met Lys
180 185 190

Lys Ile Leu Glu Met Tyr Asn Gly Phe Glu Gly Leu Thr Thr Ile Val
195 200 205

Asp Val Gly Gly Gly Thr Gly Ala Val Ala Ser Met Ile Val Ala Lys
210 215 220

Tyr Pro Ser Ile Asn Ala Ile Asn Phe Asp Leu Pro His Val Ile Gln
225 230 235 240

Asp Ala Pro Ala Phe Ser Gly Val Glu His Leu Gly Gly Asp Met Phe
245 250 255

Asp Gly Val Pro Lys Gly Asp Ala Ile Phe Ile Lys Trp Ile Cys His
260 265 270

Asp Trp Ser Asp Glu His Cys Leu Lys Leu Leu Lys Asn Cys Tyr Ala
275 280 285

Ala Leu Pro Asp His Gly Lys Val Ile Val Ala Glu Tyr Ile Leu Pro
290 295 300

Pro Ser Pro Asp Pro Ser Ile Ala Thr Lys Val Val Ile His Thr Asp
305 310 315 320

Ala Leu Met Leu Ala Tyr Asn Pro Gly Gly Lys Glu Arg Thr Glu Lys
325 330 335

Glu Phe Gln Ala Leu Ala Met Ala Ser Gly Phe Arg Gly Phe Lys Val
340 345 350

Ala Ser Cys Ala Phe Asn Thr Tyr Val Met Glu Phe Leu Lys Thr Ala
355 360 365

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION:1
(D) OTHER INFORMATION:/note= "X = any amino acid"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

Xaa Thr Gly Asn Ala Glu Thr Gln Leu Thr Pro
1 5 10

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Ser Pro Gly Asn Ala Glu Ile Gln Ile Ile Pro
1 5 10

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 amino acids
(B) TYPE: amino acid

(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

Met Leu Asp Arg Val Leu Arg Leu Leu Ala Ser Tyr Ser Val Val Thr
1 5 10 15

Tyr Thr Leu Arg Glu
20

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Met Phe Asp Gly Val Pro Lys Gly Asp Ala Ile Phe Ile Lys
1 5 10

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

Met Ala Thr Val Pro Glu Leu Asn Cys Glu Met Pro Pro Ser Asp
1 5 10 15

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2693 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

ATGGGTTCAA CAGGTGAAAC TCAGATGACT CCAACTCAGG TATCAGATGA AGAGGCACAC 60

CTCTTTGCCA TGCAACTAGC CAGTGCTTCA GTTCTACCAA TGATCCTCAA AACAGCCATC 120

GAACTCGACC TTCTTGAAAT CATGGCTAAA GCTGGCCCTG GTGCTTTCTT GTCCACATCT 180

GAGATAGCTT CTCACCTCCC TACCAAAAAC CCTGATGCGC CTGTCATGTT AGACCGTATC 240

CTGCGCCTCC TGGCTAGCTA CTCCATTCTT ACCTGCTCTC TGAAAGATCT TCCTGATGGG 300

AAGGTTGAGA GACTGTATGG CCTCGCTCCT GTTTGTAAAT TCTTGACCAA GAACGAGGAC 360

GGTGTCTCTG TCAGCCCTCT CTGTCTCATG AACCAGGGCA AGGTCCTCAT GGAAAGCTGG 420

TTAGTATCCT GTCTTCACCA ATCTAAGAAA TCCTGAATTA CATATTGAAT TTGATTATAA 480

AGTGGCTTAC AAACTCTCCA CTGAGATTTA TGTTGTTGCA CATTTGCTCT GTTTCTCAAT 540

CTTATTATGC TATAGAAAAG CAATCCAAAG TGACCAAATT GAGGGATCGG CACCACAGAC 600

TTCTCTCTCA CTAGAGACCA TTAGAGATGG GTGAATTAGG GTCCCACCAA TTTGACAATT 660

GCAAGCCACC ACTTTCCCTG CCATAAAGGT TTTGCCTGCC GGCAAATTTG TCGACCAGTC 720

CAAATGGGCA TCCCCTAAAG TTCTAGTTTT AAGAGAGAGA TATGATTAGA ATATTTTTCT 780

ACATATTTAA AGTTACTTAT GGTTAATGTC CGAAAAAATA AAAAAATGAA AACATATTGT 840

TATTGAATTT TTATAACCAT CAAACCTACC TCTCTAGGTT AGAAATTTCC TTTTCAGCTA 900

AAAGAAATTG TATTTTCCAA TGGTGATATT AACTGTTATC TAAAATAAAG TCAAATTAAT 960

ATGGTCAATT ATTGCTGTCG ATGCTTTATT TATCAAATAT GAACCTTCGA CGAAAGCATC 1020

ACTTTTTTCT CTCTCTCTCA AATTTGAGTC ATAAGGATTA ATGGATAGGC TAATTGCCAA 1080

GAATTAATTA ACTAGTTTAA ATCGACTTCT TGTAATTGCT ATGGCCATAT AGTGTTTGTC 1140

ATTAATTAAA GTGAAGGCAG TAAAAATATC TTTGCATTGC TTTTTCTTTA CTCAAATTAA 1200

AAATACTTTA ATTTGATGAT GTGTTTTTTT TAAATCGAAT AAAAAACATT TAGATAATAT 1260

TTGAATTCTG TCTCAGGATA TAAAAGACAA AAACAACATA TATAAATAAA AATATATAAA 1320

TCCATTCTAA AACTGTCCAA AAACATTTAT TTTATTTATA TTTATCTCTT CAATTAAATA 1380

TGTTATGTTT TTTTGTATTT ATCTTTTTTT GTCGTGAAAC ACTATAAAAA CGCAATGTTA 1440

AAATACAATT AATTCATTGT TTTTAGTATT CGGTGCATTT TTGAAAGCAT AATGTGATCA 1500

TGATTCTGTT AATTGTTGCA GGTATTATTT GAAAGATGCA ATTCTTGATG GAGGAATTCC 1560

ATTTAACAAG GCGTATGGGA TGACTGCATT TGAATATCAT GGCACGGATC CAAGATTCAA 1620

CAAGGTCTTC AATAAGGGAA TGTCTGACCA CTCTACCATT ACCATGAAGA AGATTCTTGA 1680

GACCTACAAA GGCTTTGAAG GCCTCACGTC CTTGGTGGAT GTTGGTGGCG GGACTGGAGC 1740

TGTCGTTAAC ACCATCGTCT CTAAATACCC TTCGATTAAG GGCATTAACT TCGATCTGCC 1800

CCATGTCATT GAGGATGCCC CATCTTATCC CGGTACTCAC GACCTTTTCT CATGTCTATT 1860

GCCAGCATTT AGATTATCTT GGTGGATATA TGCAGTAATG CTCCATTGCT TGTCCAGAAC 1920

TGCTTATGAT TTTGGACTGA ATTTGATGGA TAATGATTGC AGGAGTGGAG CATGTTGGTG 1980

GCGACATGTT TGTTAGTGTG CCCAAAGCAG ATGCCGTTTT CATGAAGGTG AGCTTTTTCC 2040

ATCCATCGGA ACCACTGTGC CACCTACTCC CTTTCATCCA TTAGCTTTTT ACTAGGACCT 2100

ACTCCTTGAA TAAACAACCA AGAGAGTCTA TATCCTTCGA TTTTGTTACT GATGTGGTGT 2160

TAATTATGCT TTTACAGTGG ATATGCCATG ATTGGAGCGA CGCCCACTGC TTAAAATTCT 2220

TGAAGAATTG CTATGACGCG TTGCCGGAAA ACGGCAAGGT GATACTTGTT GAGTGCATTC 2280

TTCCCGTGGC TCCTGACACA AGCCTCGCCA CCAAGGGAGT CGTGCACGTT GATGTTATCA 2340

TGCTGGCCCA CAACCCCGGT GGGAAAGAGA GGACCGAAAA GGAATTTGAG GGCTTAGCTA 2400

AGGGAGCTGG CTTCCAAGGT TTTGAAGTAA TGTGCTGTGC ATTCAACACA CATGTCATTG 2460

AATTCCGCAA GAAGGCCTAA GGCCCATGTC CAAGCTCCAA GTTACTTGGG GTTTTGCAGA 2520

CAACGTTGCT GCTGTCTCTG CGTTTGATGT TGTGATTGCT TTTTTTTATA CGAGGAGTAG 2580

CTATCTCTTA TGAAACATGT AAGGATAAGA TTGCGTTTTG TATGCCTGAT TTTCTCAAAT 2640

AACTTCACTG CCTCCCTCAA AATTCTTAAT ACATGTGAAA ATATTTCCTA TTG 2693

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 365 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

Met Gly Ser Thr Gly Glu Thr Gln Met Thr Pro Thr Gln Val Ser Asp
1 5 10 15

Glu Glu Ala His Leu Phe Ala Met Gln Leu Ala Ser Ala Ser Val Leu
20 25 30

Pro Met Ile Leu Lys Thr Ala Ile Glu Leu Asp Leu Leu Glu Ile Met
35 40 45

Ala Lys Ala Gly Pro Gly Ala Phe Leu Ser Thr Ser Glu Ile Ala Ser
50 55 60

His Leu Pro Thr Lys Asn Pro Asp Ala Pro Val Met Leu Asp Arg Ile
65 70 75 80

Leu Arg Leu Leu Ala Ser Tyr Ser Ile Leu Thr Cys Ser Leu Lys Asp
85 90 95

Leu Pro Asp Gly Lys Val Glu Arg Leu Tyr Gly Leu Ala Pro Val Cys
100 105 110

Lys Phe Leu Thr Lys Asn Glu Asp Gly Val Ser Val Ser Pro Leu Cys
115 120 125

Leu Met Asn Gln Gly Lys Val Leu Met Glu Ser Trp Tyr Tyr Leu Lys
130 135 140

Asp Ala Ile Leu Asp Gly Gly Ile Pro Phe Asn Lys Ala Tyr Gly Met
145 150 155 160

Thr Ala Phe Glu Tyr His Gly Thr Asp Pro Arg Phe Asn Lys Val Phe
165 170 175

Asn Lys Gly Met Ser Asp His Ser Thr Ile Thr Met Lys Lys Ile Leu
180 185 190

Glu Thr Tyr Lys Gly Phe Glu Gly Leu Thr Ser Leu Val Asp Val Gly
195 200 205

Gly Gly Thr Gly Ala Val Val Asn Thr Ile Val Ser Lys Tyr Pro Ser
210 215 220

Ile Lys Gly Ile Asn Phe Asp Leu Pro His Val Ile Glu Asp Ala Pro
225 230 235 240

Ser Tyr Pro Gly Val Glu His Val Gly Gly Asp Met Phe Val Ser Val
245 250 255

Pro Lys Ala Asp Ala Val Phe Met Lys Trp Ile Cys His Asp Trp Ser
260 265 270

Asp Ala His Cys Leu Lys Phe Leu Lys Asn Cys Tyr Asp Ala Leu Pro
275 280 285

Glu Asn Gly Lys Val Ile Leu Val Glu Cys Ile Leu Pro Val Ala Pro
290 295 300

Asp Thr Ser Leu Ala Thr Lys Gly Val Val His Val Asp Val Ile Met
305 310 315 320

Leu Ala His Asn Pro Gly Gly Lys Glu Arg Thr Glu Lys Glu Phe Glu
325 330 335

Gly Leu Ala Lys Gly Ala Gly Phe Gln Gly Phe Glu Val Met Cys Cys
340 345 350

Ala Phe Asn Thr His Val Ile Glu Phe Arg Lys Lys Ala
355 360 365

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 82 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

Val Val Thr Tyr Thr Leu Arg Glu Leu Pro Ser Gly Lys Val Glu Arg
1 5 10 15

Leu Tyr Gly Leu Ala Pro Val Cys Lys Phe Leu Thr Lys Asn Glu Asp
20 25 30

Gly Val Ser Leu Ala Pro Phe Leu Leu Thr Ala Thr Asp Lys Val Leu
35 40 45

Leu Glu Pro Trp Phe Tyr Leu Lys Asp Ala Ile Leu Glu Gly Gly Ile
50 55 60

Pro Phe Asn Lys Ala Tyr Gly Met Asn Glu Phe Asp Tyr His Gly Thr
65 70 75 80

Asp His

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 368 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

Met Gly Ser Thr Gly Asn Ala Glu Ile Gln Ile Ile Pro Thr His Ser
1 5 10 15

Ser Asp Glu Glu Ala Asn Leu Phe Ala Met Gln Leu Ala Ser Ala Ala
20 25 30

Val Leu Pro Met Ala Leu Lys Ala Ala Ile Glu Leu Asp Val Leu Glu
35 40 45

Ile Met Ala Lys Ser Val Pro Pro Ser Gly Tyr Ile Ser Pro Ala Glu
50 55 60

Ile Ala Ser Gln Leu Pro Thr Thr Asn Pro Glu Ala Pro Val Met Leu
65 70 75 80

Asp Arg Val Leu Arg Leu Leu Ala Ser Tyr Ser Val Val Thr Tyr Thr
85 90 95

Leu Arg Glu Leu Pro Ser Gly Lys Val Glu Arg Leu Tyr Gly Leu Ala
100 105 110

Pro Val Cys Lys Phe Leu Thr Lys Asn Glu Asp Gly Val Ser Leu Ala
115 120 125

Pro Phe Leu Leu Thr Ala Thr Asp Lys Val Leu Leu Glu Pro Trp Phe
130 135 140

Tyr Leu Lys Asp Ala Ile Leu Glu Gly Gly Ile Pro Phe Asn Lys Ala
145 150 155 160

Tyr Gly Met Asn Glu Phe Asp Tyr His Gly Thr Asp His Arg Phe Asn
165 170 175

Lys Val Phe Asn Lys Gly Met Ser Ser Asn Ser Thr Ile Thr Met Lys
180 185 190

Lys Ile Leu Glu Met Tyr Asn Gly Phe Glu Gly Leu Thr Thr Ile Val
195 200 205

Asp Val Gly Gly Gly Thr Gly Ala Val Ala Ser Met Ile Val Ala Lys
210 215 220

Tyr Pro Ser Ile Asn Ala Ile Asn Phe Asp Leu Pro His Val Ile Gln
225 230 235 240

Asp Ala Pro Ala Phe Ser Gly Val Glu His Leu Gly Gly Asp Met Phe
245 250 255

Asp Gly Val Pro Lys Gly Asp Ala Ile Phe Ile Lys Trp Ile Cys His
260 265 270

Asp Trp Ser Asp Glu His Cys Leu Lys Leu Leu Lys Asn Cys Tyr Ala
275 280 285

Ala Leu Pro Asp His Gly Lys Val Ile Val Ala Glu Tyr Ile Leu Pro
290 295 300

Pro Ser Pro Asp Pro Ser Ile Ala Thr Lys Val Val Ile His Thr Asp
305 310 315 320

Ala Leu Met Leu Ala Tyr Asn Pro Gly Gly Lys Glu Arg Thr Glu Lys
325 330 335

Glu Phe Gln Ala Leu Ala Met Ala Ser Gly Phe Arg Gly Phe Lys Val
340 345 350

Ala Ser Cys Ala Phe Asn Thr Tyr Val Met Glu Phe Leu Lys Thr Ala
355 360 365

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 370 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

Met Gly Ser Thr Gly Asn Ala Glu Thr Gln Leu Thr Pro Thr His Val
1 5 10 15

Ser Asp Glu Glu Ala Asn Leu Phe Ala Met Gln Leu Ala Ser Ala Ser
20 25 30

Val Leu Pro Met Val Leu Lys Ala Ala Ile Glu Leu Asp Val Leu Glu
35 40 45

Ile Met Ala Lys Ser Ile Pro His Gly Ser Gly Ala Tyr Ile Ser Pro
50 55 60

Ala Glu Ile Ala Ala Gln Leu Pro Thr Thr Asn Pro Asp Ala Pro Val
65 70 75 80

Met Leu Asp Arg Val Leu Arg Leu Leu Ala Ser Tyr Ser Val Val Thr
85 90 95

Cys Ser Leu Arg Glu Leu Pro Asp Gly Lys Val Glu Arg Leu Tyr Gly
100 105 110

Leu Ala Pro Val Cys Lys Phe Leu Thr Lys Asn Glu Asp Gly Val Ser
115 120 125

Leu Ala Pro Leu Cys Leu Met Asn Gln Asp Lys Val Leu Met Glu Ser
130 135 140

Trp Tyr Tyr Leu Lys Asp Ala Ile Leu Asp Gly Gly Ile Pro Phe Asn
145 150 155 160

Lys Ala Tyr Gly Met Thr Ala Phe Glu Tyr His Gly Thr Asp Pro Arg
165 170 175

Phe Asn Lys Val Phe Asn Arg Gly Met Ser Asp His Ser Thr Ile Thr
180 185 190

Met Lys Lys Ile Phe Glu Met Tyr Thr Gly Phe Glu Ala Leu Asn Thr
195 200 205

Ile Val Asp Val Gly Gly Gly Thr Gly Ala Val Leu Ser Met Ile Val
210 215 220

Ala Lys Tyr Pro Ser Ile Lys Gly Ile Asn Phe Asp Leu Pro His Val
225 230 235 240

```
Ile Glu Asp Ala Pro Ile Tyr Pro Gly Val Glu His Val Gly Gly Asp
                245                 250                 255

Met Phe Val Ser Val Pro Lys Gly Asp Ala Ile Phe Met Lys Trp Ile
            260                 265                 270

Cys His Asp Trp Ser Asp Glu His Cys Leu Lys Phe Leu Lys Asn Cys
        275                 280                 285

Tyr Ala Ala Leu Pro Glu His Gly Lys Val Ile Val Ala Glu Cys Ile
    290                 295                 300

Leu Pro Leu Ser Pro Asp Pro Ser Leu Ala Thr Lys Gly Val Ile His
305                 310                 315                 320

Ile Asp Ala Ile Met Leu Ala His Asn Pro Gly Gly Lys Glu Arg Thr
                325                 330                 335

Glu Lys Glu Phe Glu Ala Leu Ala Ile Gly Ala Gly Phe Lys Gly Phe
            340                 345                 350

Lys Val Ala Cys Cys Ala Phe Asn Thr Tyr Val Met Glu Phe Leu Lys
        355                 360                 365

Thr Ala
    370
```

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 369 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

```
Met Gly Ser Thr Gly Glu Thr Gln Met Thr Pro Thr Gln Val Ser Asp
1               5                   10                  15

Glu Glu Ala His Leu Phe Ala Met Gln Leu Ala Ser Ala Ser Val Leu
            20                  25                  30

Pro Met Ile Leu Lys Thr Ala Ile Glu Leu Asp Leu Leu Glu Ile Met
        35                  40                  45

Ala Lys Ala Gly Pro Gly Ala Phe Leu Ser Thr Ser Glu Ile Ala Ser
    50                  55                  60

His Leu Pro Thr Lys Asn Pro Asp Ala Pro Val Met Leu Asp Arg Ile
65                  70                  75                  80

Leu Arg Leu Leu Ala Ser Tyr Ser Ile Leu Thr Cys Ser Leu Lys Asp
                85                  90                  95
```

Leu Pro Asp Gly Lys Val Glu Arg Leu Tyr Gly Leu Ala Pro Val Cys
100 105 110

Lys Phe Leu Thr Lys Asn Glu Asp Gly Val Ser Val Ser Pro Leu Cys
115 120 125

Leu Met Asn Gln Asp Lys Val Leu Met Glu Ser Trp Tyr Tyr Leu Lys
130 135 140

Asp Ala Ile Leu Asp Gly Gly Ile Pro Phe Asn Lys Ala Tyr Gly Met
145 150 155 160

Thr Ala Phe Glu Tyr His Gly Thr Asp Pro Arg Phe Asn Lys Val Phe
165 170 175

Asn Lys Gly Met Ser Asp His Ser Thr Ile Thr Met Lys Lys Ile Leu
180 185 190

Glu Thr Tyr Lys Gly Phe Glu Gly Leu Thr Ser Leu Val Asp Val Gly
195 200 205

Gly Gly Thr Gly Ala Val Val Asn Thr Ile Val Ser Lys Tyr Pro Ser
210 215 220

Ile Lys Gly Ile Asn Phe Asp Leu Pro His Val Ile Glu Asp Ala Pro
225 230 235 240

Ser Tyr Pro Gly Val Glu His Val Gly Gly Asp Met Phe Val Ser Val
245 250 255

Pro Lys Ala Asp Ala Val Phe Met Lys Trp Ile Cys His Asp Trp Ser
260 265 270

Asp Ala His Cys Leu Lys Phe Leu Lys Asn Cys Tyr Asp Ala Leu Pro
275 280 285

Glu Asn Gly Lys Val Ile Leu Val Glu Cys Ile Leu Pro Val Ala Pro
290 295 300

Asp Thr Ser Leu Ala Thr Lys Gly Val Val His Val Asp Val Ile Met
305 310 315 320

Leu Ala His Asn Pro Gly Gly Lys Glu Arg Thr Glu Lys Glu Phe Glu
325 330 335

Gly Leu Ala Lys Gly Ala Gly Phe Gln Gly Phe Glu Val Met Cys Cys
340 345 350

Ala Phe Asn Thr His Val Ile Glu Phe Arg Lys Lys Ala Ala Cys Asp
355 360 365

Cys

CLAIMS:

1. An isolated nucleic acid comprising a nucleic acid segment coding for (iso)eugenol methyl transferase (IEMT) or an active fragment thereof.

2. The isolated nucleic acid of claim 1, wherein said nucleic acid segment encodes full length IEMT.

3. The isolated nucleic acid of claim 2, wherein said IEMT has the sequence of SEQ ID NO:2.

4. The isolated nucleic acid of claim 3, wherein said nucleic acid segment has the sequence of SEQ ID NO:1.

5. An expression vector comprising a nucleic acid segment coding for (iso)eugenol methyl transferase (IEMT) and a promoter operatively linked to said nucleic acid segment.

6. The expression vector of claim 5, wherein said promoter is selected from the group consisting of 35S promoter and CHS-A promoter.

7. The expression vector of claim 5, wherein said promoter is a fruit-specific promoter or a leaf-specific promoter.

8. The expression vector of claim 5, further comprising an origin of replication and a polyadenylation signal.

9. A method for increasing the synthesis of methyl-isoeugenol in a plant cell comprising the steps of:

(i) providing plant cells;

(ii) contacting said plant cell with a nucleic acid segment coding for (iso)eugenol methyl transferase (IEMT) and a promoter operatively linked to said nucleic acid segment; and

(iii) selecting a cell having said nucleic acid stably integrated into is genome.

10. The method of claim 9, wherein said plant cell is a monocot.

11. The method of claim 9, wherein said plant cell is a dicot.

12. The method of claim 11, wherein said dicot plant cell is a tomato cell.

13. The method of claim 9, wherein said contacting comprises microprojectile bombardment, electroporation or *Agrobacterium*-mediated transformation.

14. The method of claim 9, wherein said selecting comprises identifying a cell having drug resistance.

15. The method of claim 14, wherein said drug resistance is selected from the group consisting of kanamycin and hygromycin.

16. A transgenic plant cell having, incorporated into its genome, a nucleic acid segment coding for (iso)eugenol methyl transferase (IEMT) and a promoter operatively linked to said nucleic acid segment.

17. The transgenic plant cell of claim 16, wherein said plant cell is a monocot.

18. The transgenic plant cell of claim 16, wherein said plant cell is a dicot.

19. The transgenic plant cell of claim 18, wherein said dicot cell is a tomato.

20. A transgenic plant having, incorporated into the genome of cells of said plant, a nucleic acid segment coding for (iso)eugenol methyl transferase (IEMT) and a promoter operatively linked to said nucleic acid segment.

21. The transgenic plant of claim 20, wherein said plant is a monocot.

22. The transgenic plant of claim 20, wherein said plant is a dicot.

23. The transgenic plant of claim 22, wherein said dicot is a tomato.

24. A method for decreasing the (iso)eugenol content of a plant cell comprising the steps of:

(i) providing plant cells;

(ii) contacting said plant cell with a nucleic acid segment coding for (iso)eugenol methyl transferase (IEMT) and a promoter operatively linked to said nucleic acid segment; and

(iii) selecting a cell having said nucleic acid stably integrated into its genome.

25. The method of claim 24, wherein said contacting comprises microprojectile bombardment, electroporation or Agrobacterium-mediated transformation.

26. A method for decreasing lignin biosynthesis in a plant cell comprising the steps of:

(i) providing a plant cell;

(ii) contacting said plant cell with nucleic acid segement encoding a chimeric enzyme having amino acid sequences from (iso)eugenol methyl transferase (IEMT) and coumaric acid methyl transferase (COMT), wherein the substrate specificity of said enzyme includes *p*-coumaric acid, caffeic acid, ferulic acid, eugenol and isoeugenol and the activity comprises 4-hydroxyl methylation; and

(iii) selecting a cell having said nucleic acid segment stably integrated into its genome.

27. The method of claim 26, wherein said contacting comprises microprojectile bombardment, electroporation or Agrobacterium-mediated transformation.

28. A transgenic plant having, incorporated into the genome of cells of said plant, a nucleic acid segment encoding a chimeric enzyme having amino acid sequences from (iso)eugenol methyl transferase (IEMT) and caffeic acid methyl transferase (COMT), wherein the substrate specificity of said enzyme includes *p*-coumaric acid, caffeic acid, ferulic acid, eugenol and isoeugenol and the activity comprises 4-hydroxyl methylation and a promoter operatively linked to said nucleic acid segment.

29. An isolated nucleic acid segment encoding a chimeric enzyme having amino acid sequences from (iso)eugenol methyl transferase (IEMT) and caffeic acid methyl transferase (COMT), wherein the substrate specificity of said enzyme includes *p*-coumaric acid, caffeic acid, ferulic acid, eugenol and isoeugenol and the activity comprises 4-hydroxyl methylation and a promoter operatively linked to said nucleic acid segment.

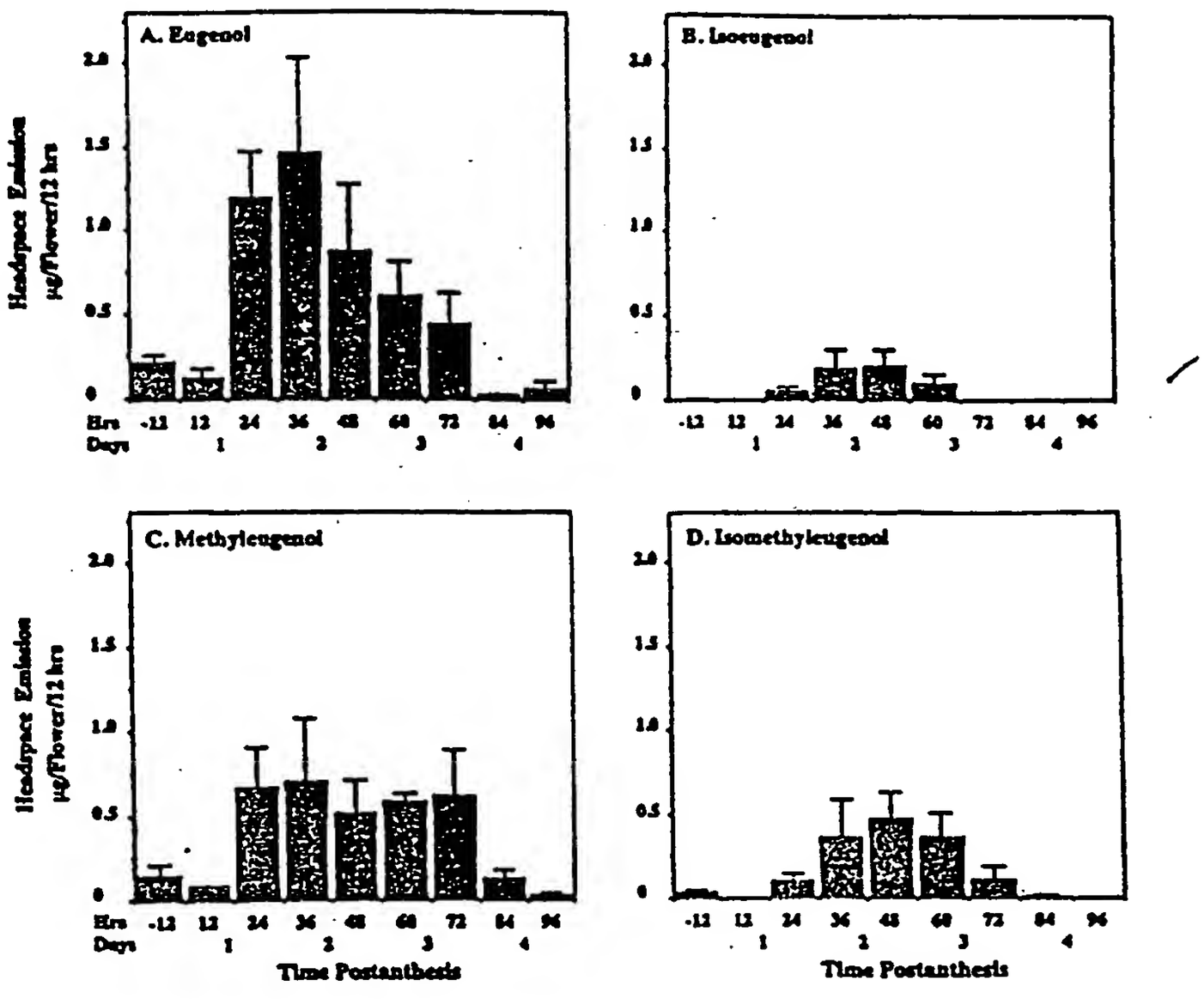

A. Eugenol
B. Isoeugenol
C. Methyleugenol
D. Isomethyleugenol
Headspace Emission µg/flower/12 hrs
2.0
1.5
1.0
0.5
0
Hrs -12 12 24 36 48 60 72 84 96
Days 1 2 3 4
Time Postanthesis

FIG. 1A - FIG. 1D

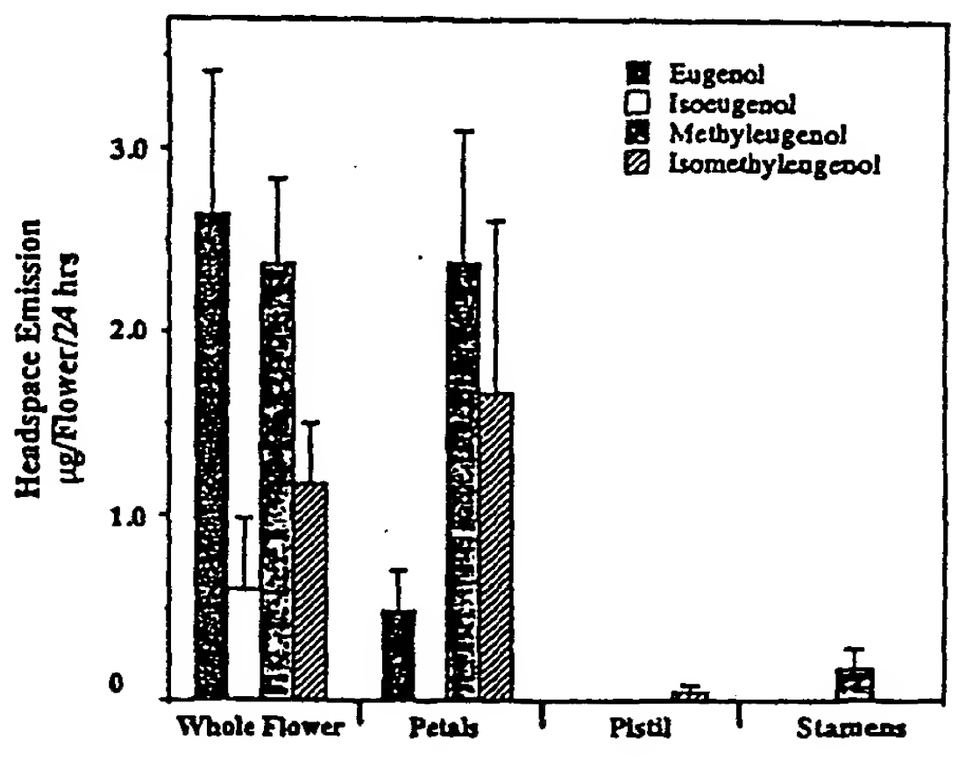

Eugenol
Isoeugenol
Methyleugenol
Isomethyleugenol
Headspace Emission
µg/Flower/24 hrs
3.0
2.0
1.0
0
Whole Flower
Petals
Pistil
Stamens

Fig. 2

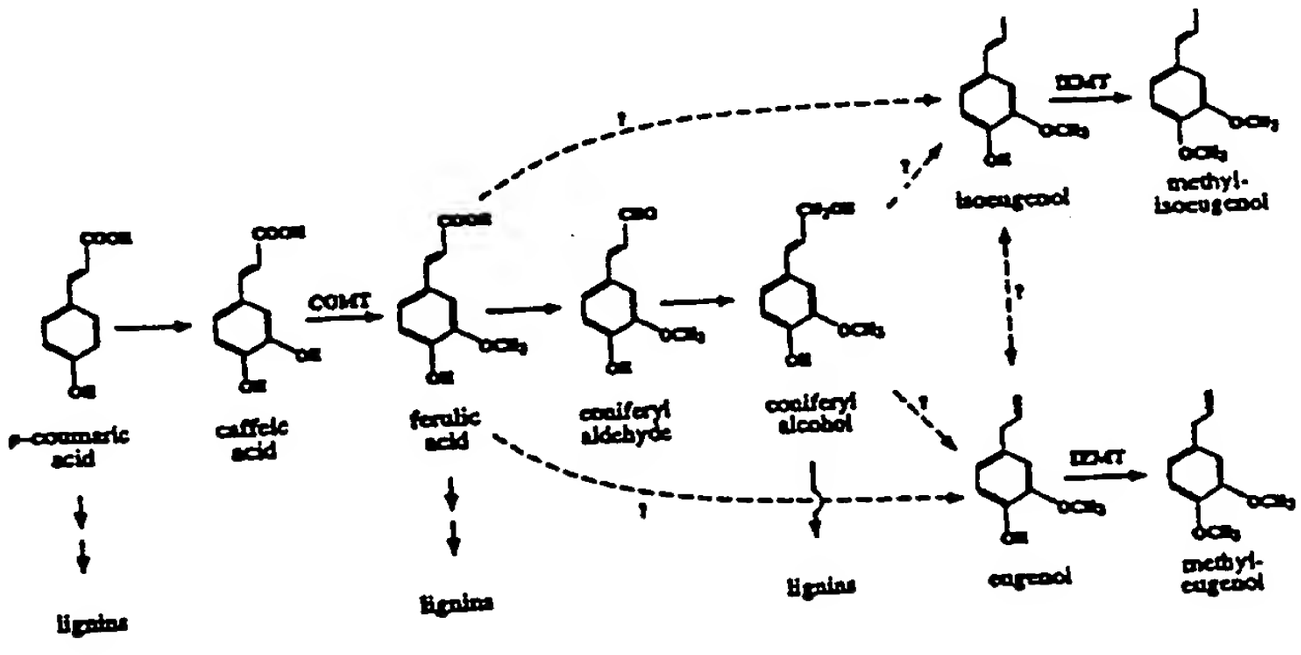

COMT
CCOMT
?
p-coumaric acid
caffeic acid
ferulic acid
coniferyl aldehyde
coniferyl alcohol
isoeugenol
methyl-isoeugenol
eugenol
methyl-eugenol
lignins
lignins
lignins

Fig. 3

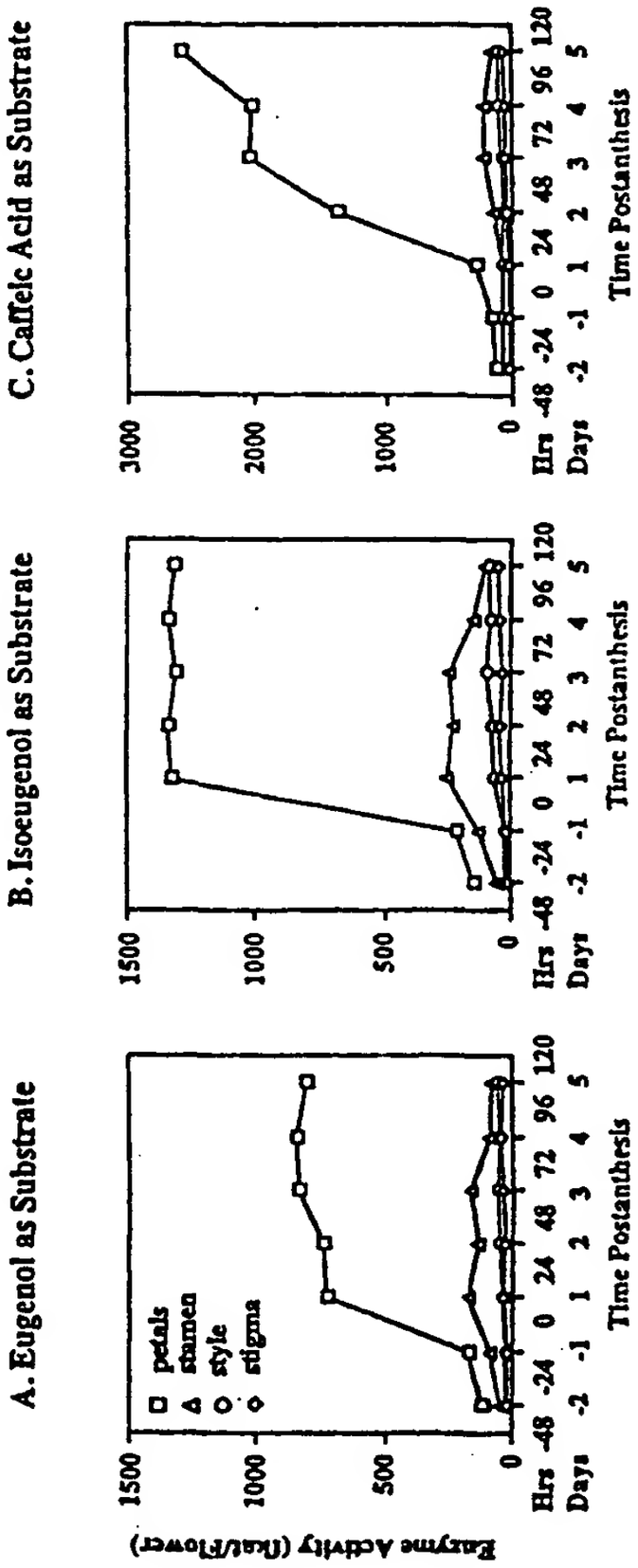


Fig. 4A

Fig. 4B

Fig. 4C

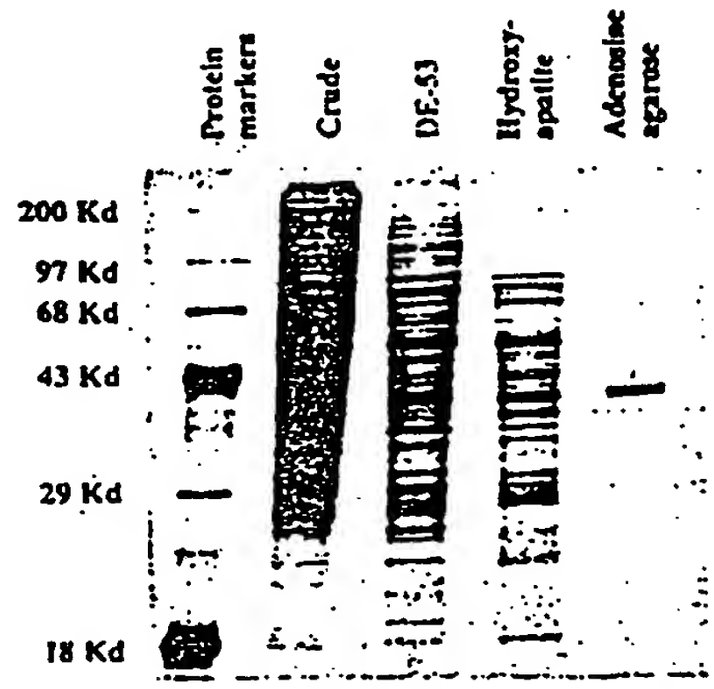

Protein markers
Crude
DE-53
Hydroxy-apatite
Adenosine agarose
200 Kd
97 Kd
68 Kd
43 Kd
29 Kd
18 Kd

Fig. 5

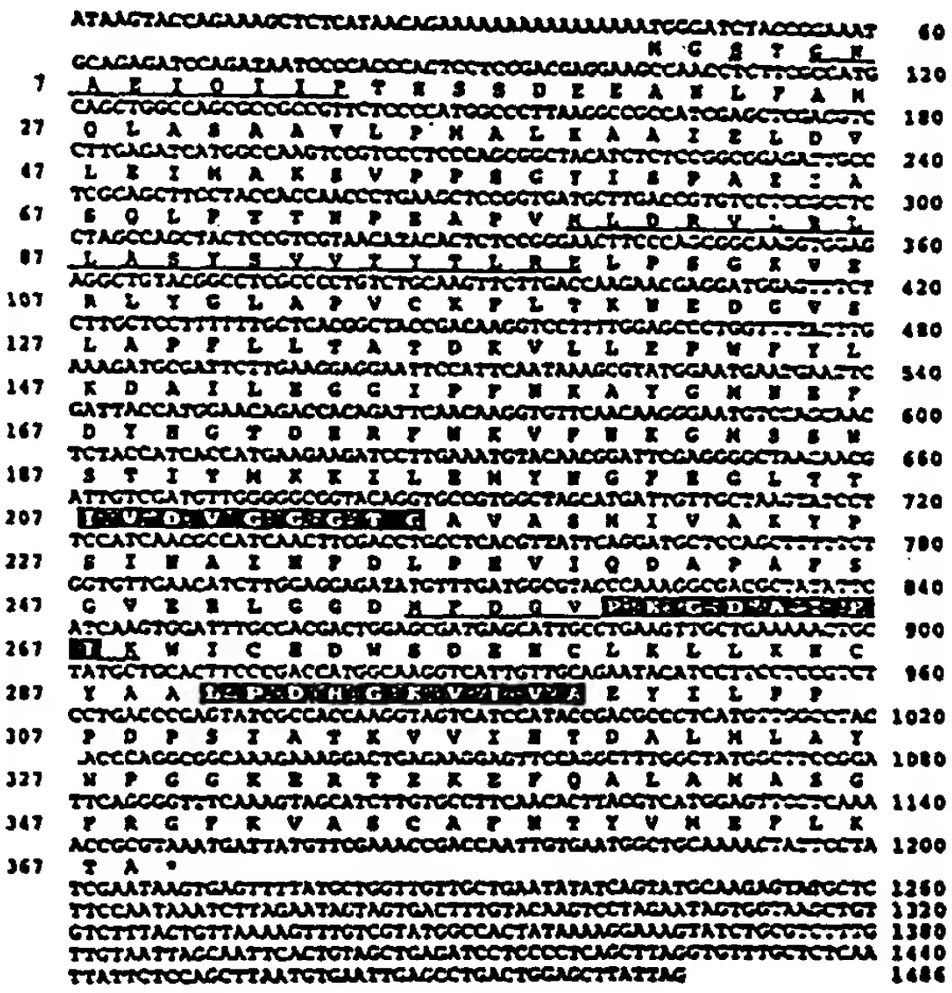
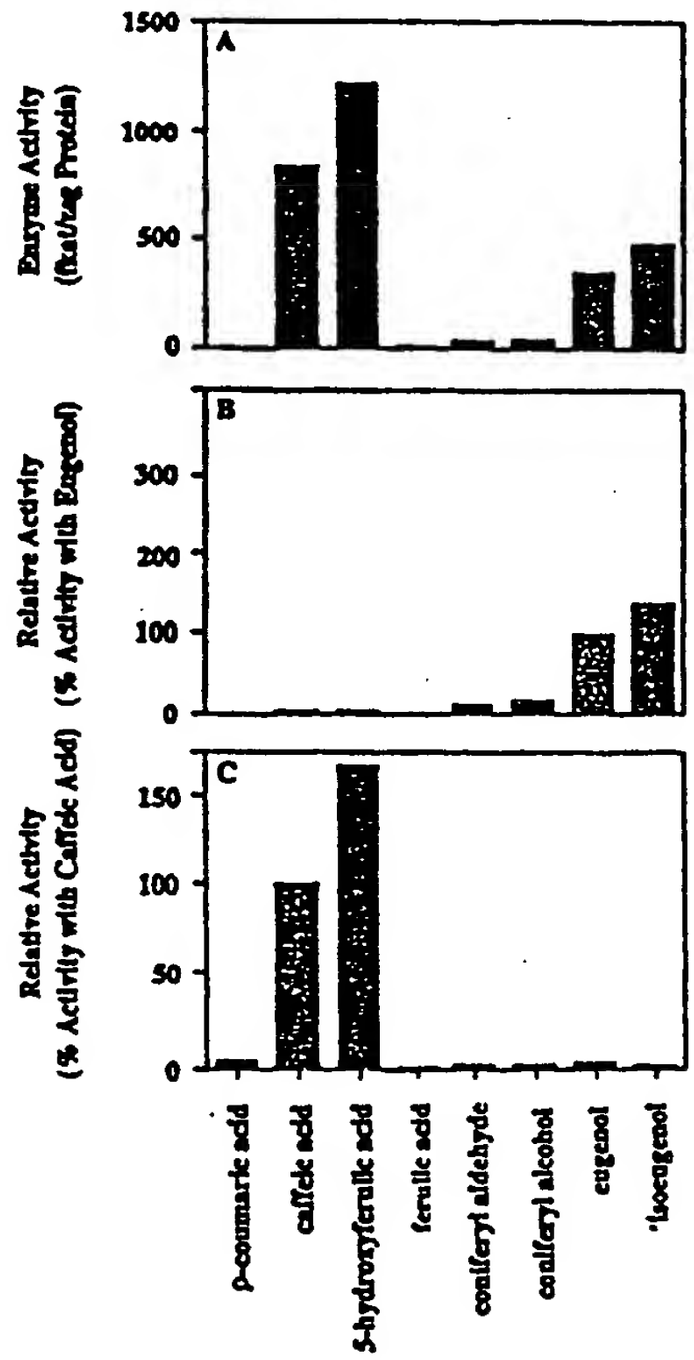


Fig. 6

Fig. 7A

Fig. 7B

Fig. 7C



A (Iso)Methyleugenol-Emitting Plants

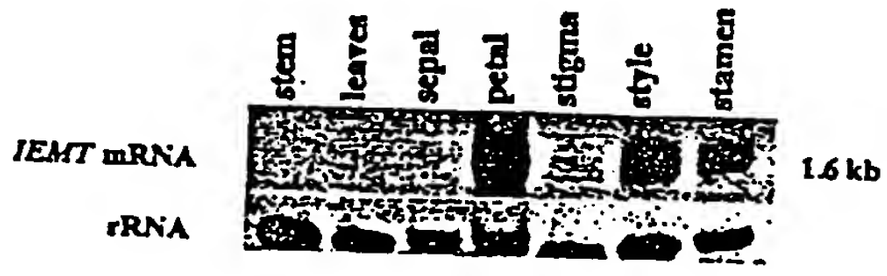


FIG. 8A

B (Iso)Methyleugenol-Nonemitting Plants

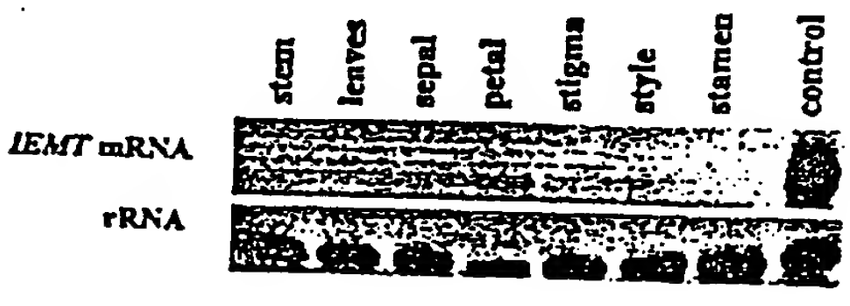


FIG. 8B

A

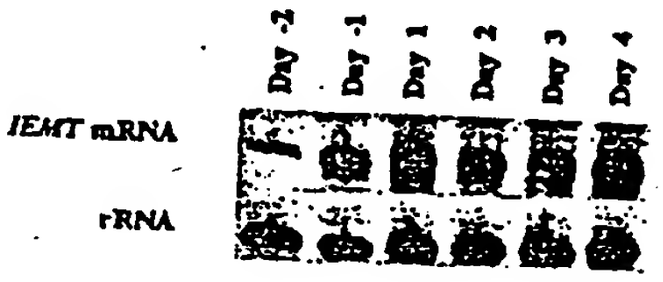


Fig. 9A

B

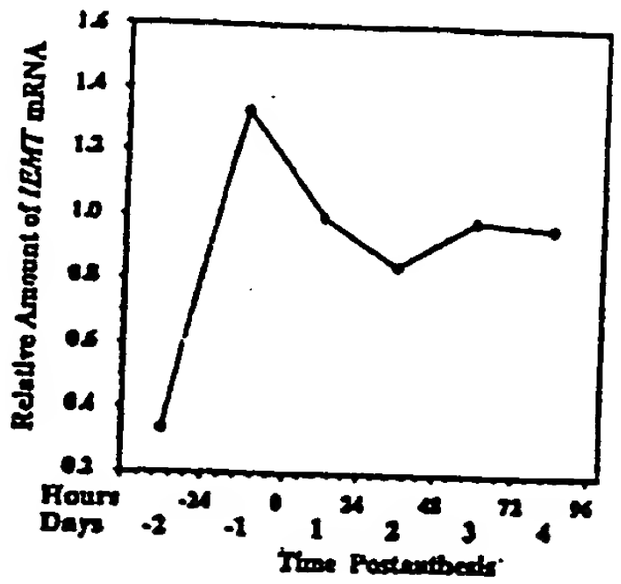


Fig. 9B

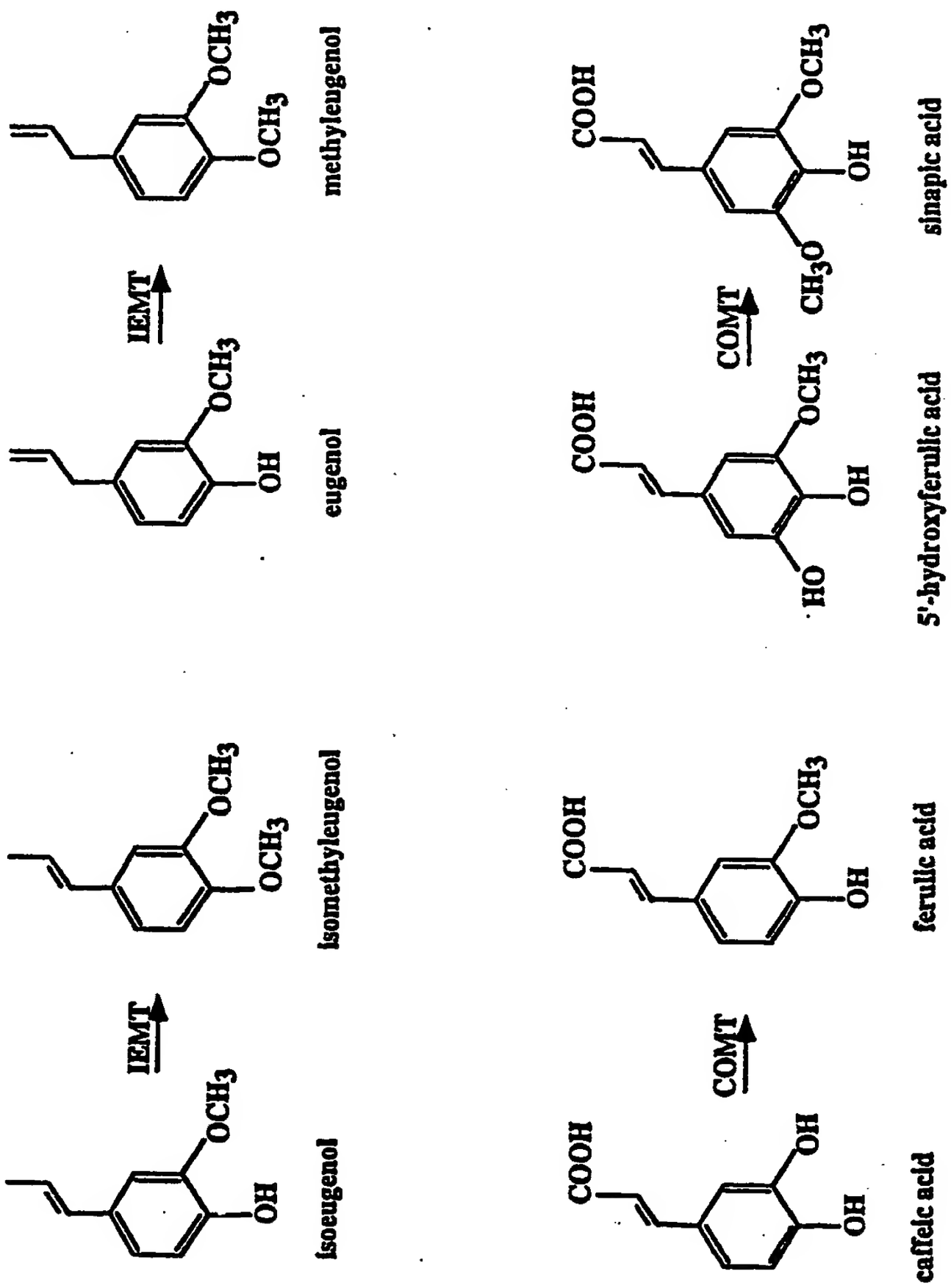


FIG. 10

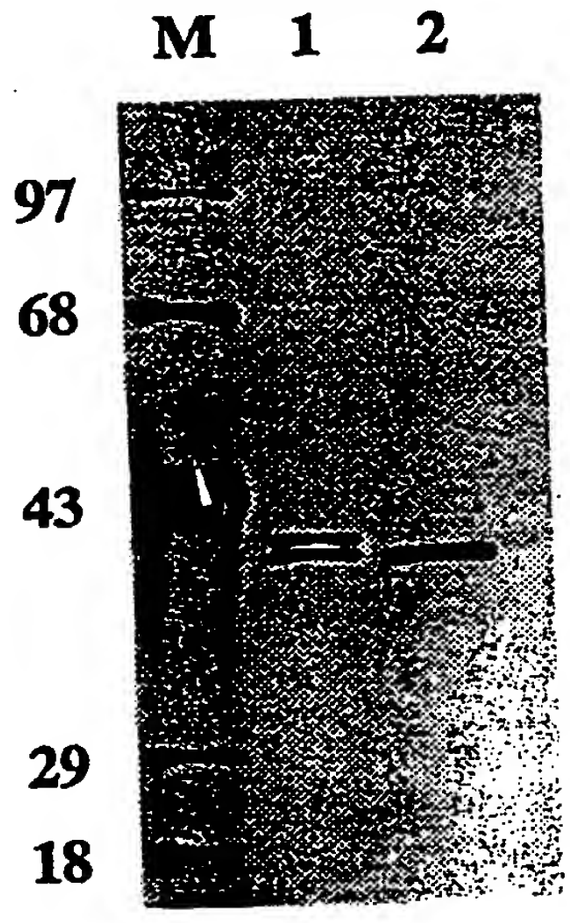


FIG. 11

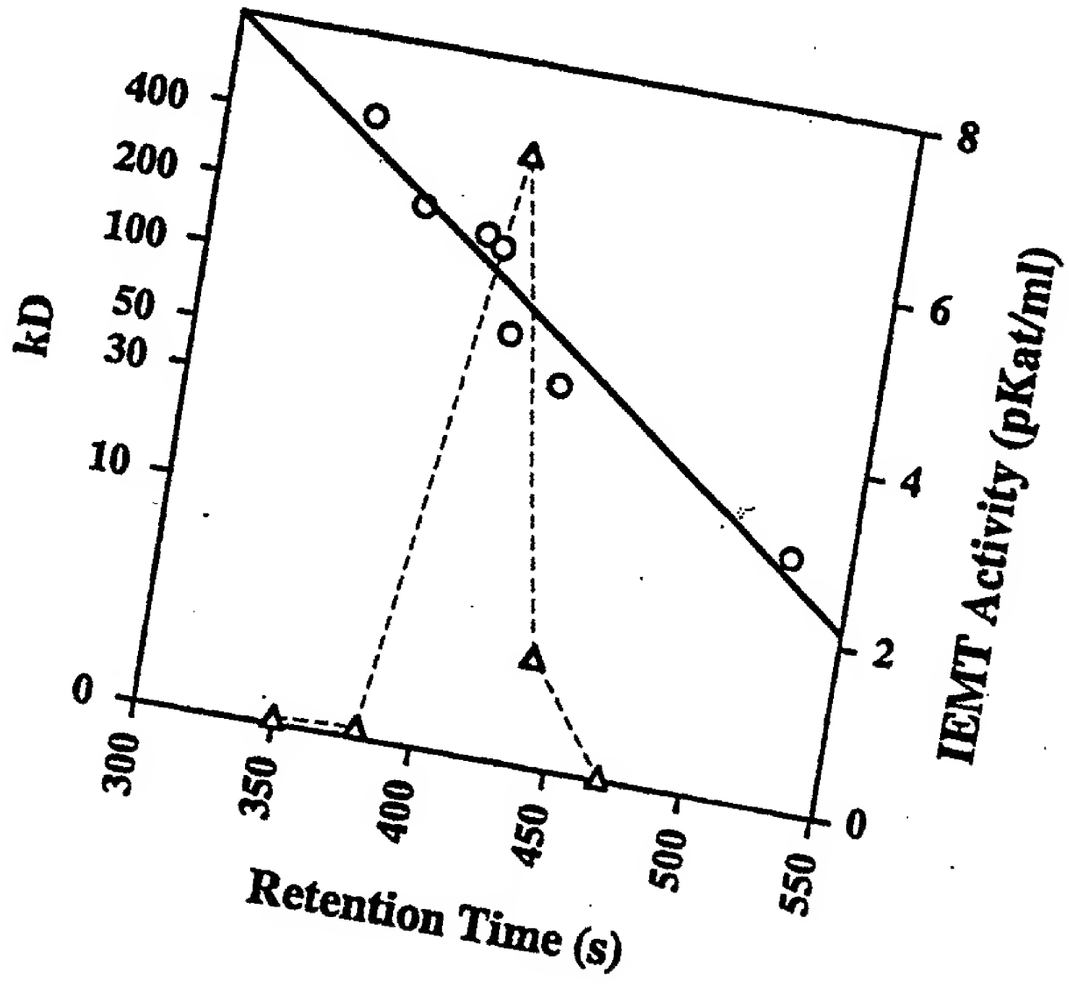

400
200
100
50
30
10
0
kD
300
350
400
450
500
550
Retention Time (s)
8
6
4
2
0
IEMT Activity (pKat/ml)

FIG. 12

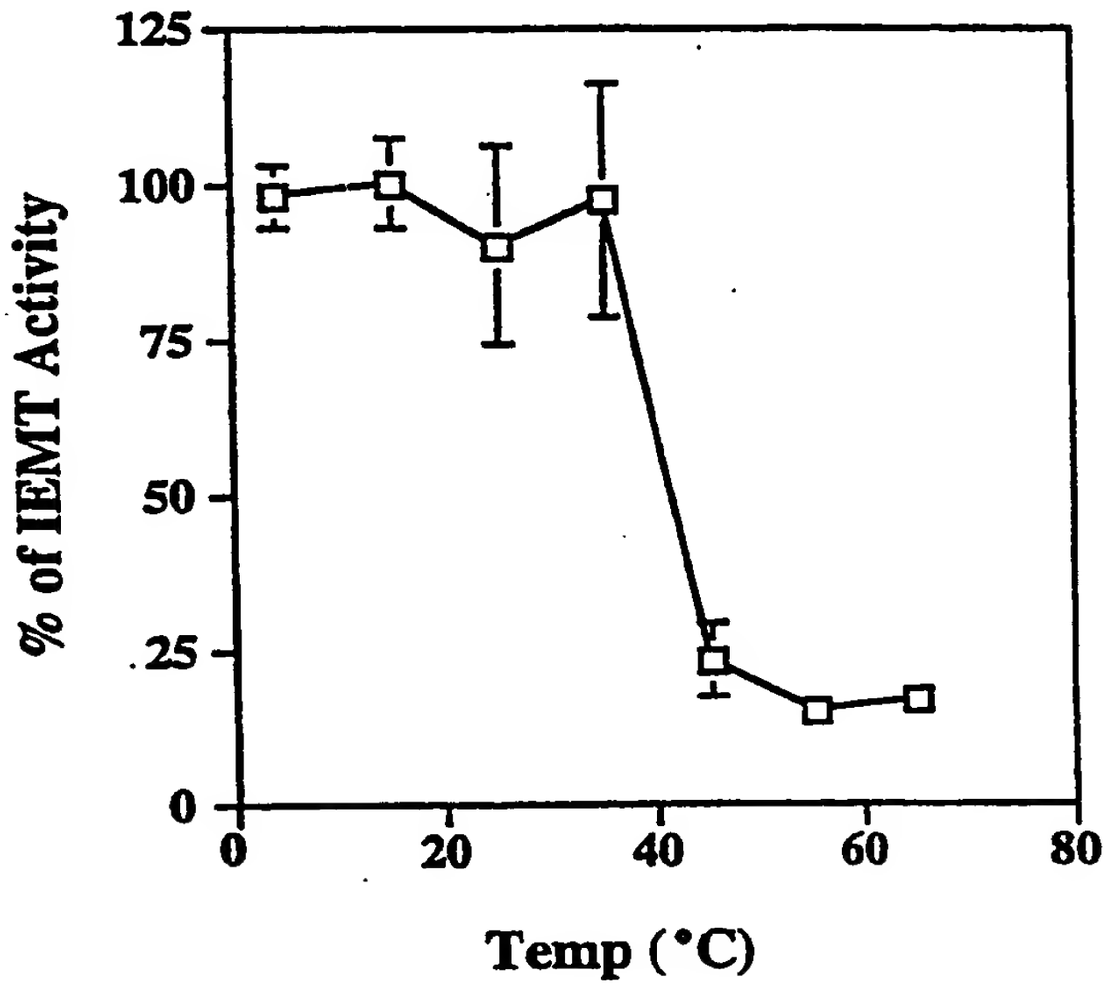


FIG. 13

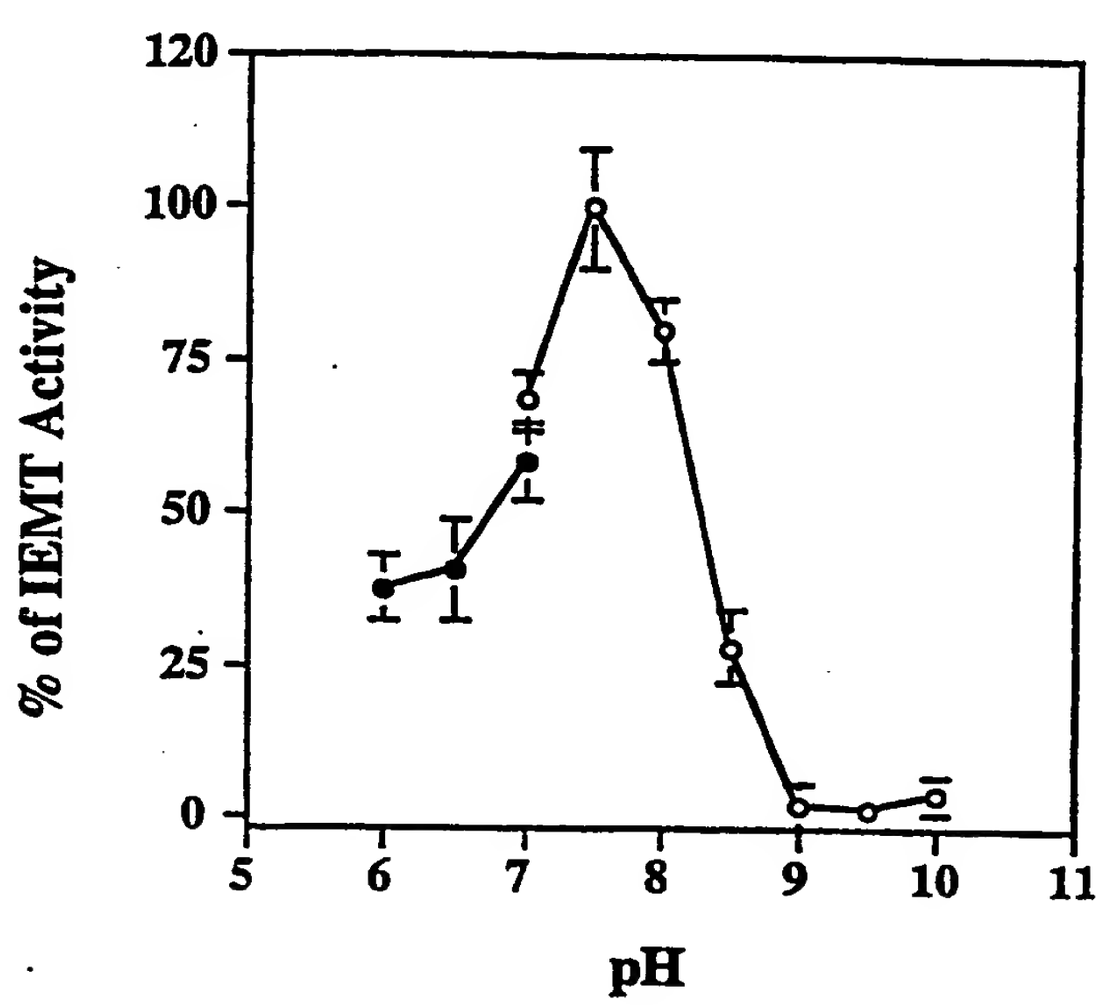


FIG. 14

```
                                                                  46
IEMT-CB: MGSTGNAEIQIIPTHSSDEEANLFAMQLASAAVLPMALKAAIELDVLEIM     50
COMT-CB: ........T.LT...V.................S....V.............
COMT-PT: .....--.T.MT..QV.....H.........S....I..T.....L....

                                                                  92
IEMT-CB: AKS--VPPSGYISPAEIASKLPTTNPEAPVMLDRVLRLLASYSVVTYTLREL 100
COMT-CB: ...IPHGSGA.........AQ......D.....................CS....
COMT-PT: ..A---G.GAFL..S....H...K..D.......I........IL.CS.KD.

IEMT-CB: PSGKVERLYGLAPVCKFLTKNEDGVSLAPFLLTATDKVLLEPWFYLKDAI    150
COMT-CB: .D.............................LC.MNQ....M.S.Y......
COMT-PT: .D.........................VS.LC.MNQ....M.S.Y......

                                                  173
IEMT-CB: LEGGIPFNKAYGMNEFDYHGTDHRFNKVFNKGMSSNSTITMKKILEMYNG    200
COMT-CB: .D...........TA.E.....P.......R...DH........F...T.
COMT-PT: .D...........TA.E.....P...........DH..........T.K.

                  207
IEMT-CB: FEGLTTIVDVGGGTGAVASMIVAKYPSINAINFDLPHVIQDAPAFSGVEH    250
COMT-CB: ..A.N.............L...........KG.........E...IYP....
COMT-PT: .....SL.........VNT..S.....KG.........E...SYP....

IEMT-CB: LGGDMFDGVPKGDAIFIKWICHDWSDEHCLKLLKNCYAALPDHGKVIVAE    300
COMT-CB: V.....VS.........M..............F.........E........
COMT-PT: V.....VS...A..V.M.........A....F.....D...EN....LV.

IEMT-CB: YILPPSPDPSIATKVVIHTDALMLAYNPGGKERTEKEFQALAMASGFRGF    350
COMT-CB: C...L.....L...G...I..I...H............E...IGA..K..
COMT-PT: C...VA..T.L...G.V.V.V....H............EG..KGA..Q..

IEMT-CB: KVASCAFNTYVMEFLKTA                                   368
COMT-CB: ...C..............
COMT-PT: E.MC.....H.I..R.K.
```

FIG. 15

Name	Construct	Substrate	K_m (μM)	K_{cat} (S^{-1})	K_{cat}/K_m ($nM^{-1}S^{-1}$)
IEMT - Cb	1 — 368	isoeugenol	74	0.08	1.12
COMT - Pt	1 — 368	caffeic acid	40	0.40	9.93
I/C 92-207	92 — 207	caffeic acid	30	0.20	6.53
I/C 92-173	92 — 173	caffeic acid	155	0.26	1.70
I/C 46-83	46 — 91	isoeugenol	65	0.02	0.32
I/C 173-207	173 — 207	isoeugenol	64	0.10	1.61

FIG. 16

FIG. 17

```
           92                                                                               173   IEMT/COMT
IEMT-CB:   VVTYTLRELPSGKVERLYGLAPVCKFLTKNEDGVSLAPFLLTATDKVLLEPWFYLKDAILEGGIPFNKAYGMNEFDYHGTDH    33.40
COMT-CB:   ...CS.....D...........................LC.MNQ....M.S.Y.......D...........TA.E.....P     0.03
COMT-PT:   IL.CS.KD..D........................VS.LC.MNQ....M.S.Y.......D...........TA.E.....P     0.04

     1:    .................................................................................P    10.54
     2:    ..........D.......................................................................    13.10
     3:    ........................................................................TA........     3.18
     4:    .........................................MNQ......................................     2.98
     5:    .........................................MNQ............................TA........     0.13
     6:    ......................................LC..........................................     9.38
     7:    ......................................LC................................TA........     0.98
     8:    ......................................LC.MNQ......................................     0.05
     9:    ......................................LC.MNQ............................TA........     0.02
```

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classificati n 6 :
C12N 15/82, 15/54, 9/10, 5/10, A01H 5/00

A3

(11) International Publication Number: **WO 98/50570**

(43) International Publication Date: 12 November 1998 (12.11.98)

(21) International Application Number: PCT/US98/09522

(22) International Filing Date: 8 May 1998 (08.05.98)

(30) Priority Data:
60/046,857 8 May 1997 (08.05.97) US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application
US 60/046,857 (CIP)
Filed on 8 May 1997 (08.05.97)

(71) Applicant *(for all designated States except US)*: THE REGENTS OF THE UNIVERSITY OF MICHIGAN [US/US]; Room 2071, Wolverine Tower, 3003 South State Street, Ann Arbor, MI 48109–1280 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: PICHERSKY, Eran [US/US]; 325 Burlington Woods Drive, No. 180, Ann Arbor, MI 48104 (US). WANG, Jihong [CN/US]; Apartment 21, 1738 Murfin, Ann Arbor, MI 48105 (US).

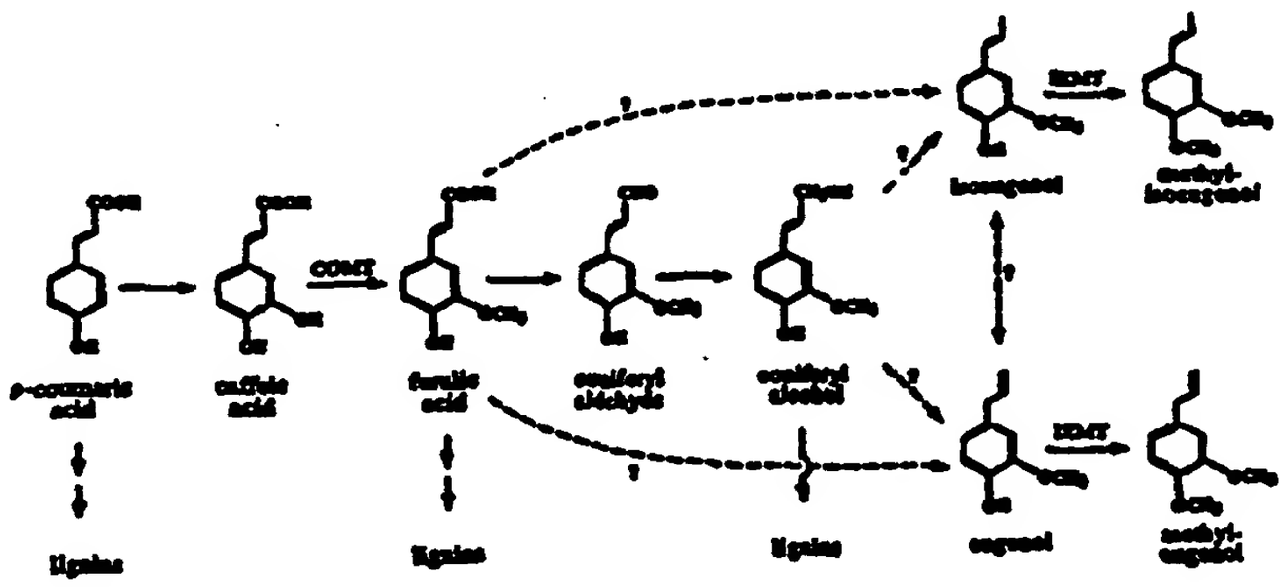
(74) Agent: MCMILLIAN, Nabeela, R.; Arnold, White & Durkee, P.O. Box 4433, Houston, TX 77210 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(88) Date of publication of the international search report:
4 February 1999 (04.02.99)

(54) Title: METHODS AND COMPOSITIONS FOR USE OF (ISO)EUGENOL METHYLTRANSFERASE



(57) Abstract

The present invention concerns DNA compositions and their use in manipulating the biosynthesis of compound in plants. More specifically, the present invention has identified a novel gene and protein involved in the phenylpropanoid pathways. *S*–adenosyl–L–Met:(iso)eugenol *O*–methyltransferase (IEMT) catalyzes a methyl transfer to the C4 hydroxyl position of compounds of the phenylpropanoid biosynthetic pathway. Methods and compositions for the use thereof are disclosed herein.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 98/09522

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/82 C12N15/54 C12N9/10 C12N5/10 A01H5/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WANG,J., ET AL. : "floral scent production in Clarkia breweri (Onagraceae)" PLANT PHYSIOLOGY, vol. 114, May 1997, pages 213-221, XP002083951 see the whole document ---	1-5
A	WO 97 15584 A (UNIV MICHIGAN) 1 May 1997 pages 3,6,19-23, page 25-27; examples, claims ---	1-11
A	WO 93 05160 A (ICI PLC) 18 March 1993 pages 4-6; pages 7-8; examples ---	1-11

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
18 November 1998	04/12/1998

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Holtorf, S

2

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

'national Application No
PCT/US 98/09522

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document. with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	WEITING, N.L., ET AL: "REDUCED LIGNIN IN TRANSGENIC PLANTS CONTAINING A CAFFEIC ACID O-METHYLTRANSFERASE ANTISENSE GENE" TRANSGENIC RESEARCH, vol. 3, no. 2, March 1994, pages 120-126, XP002054140 see the whole document	1-11
P,X	WANG,J. AND PICHERSKY, E.: "characterization of S-adenosyl-L-methionine:(iso)eugenol O-methyltransferase involved in floral scent production in Clarkia breweri" ARCHIVES OF BIOCHEMISTRY AND BIOPHYSICS, vol. 349, no. 1, 1998, pages 153-160, XP002083952 see the whole document	1-5,29
P,X	WANG,J., ET AL. : "expression of genes involved in floral scent production in Clarkia" PLANT PHYSIOLOGY, SUPPLEMENT, vol. 114, no. 3, July 1997, page 232 XP002083953 see the whole document	1-4

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 98/09522

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9715584 A	01-05-1997	AU 7663296 A	15-05-1997
		EP 0858507 A	19-08-1998
WO 9305160 A	18-03-1993	AU 663726 B	19-10-1995
		AU 2516792 A	05-04-1993
		BR 9206481 A	31-10-1995
		EP 0603250 A	29-06-1994
		JP 6510429 T	24-11-1994